TNO-report
TM-96-A037

TNO Human Factors
Research Institute

Kampweg 5
P.O. Box 23
3769 ZG Soesterberg
The Netherlands

Phone +31 346 35 62 11
Fax +31 346 35 39 77

title

Acquisition of sea targets. Part 1: Observer performance and "ACQUIRE" model predictions for air-to-surface FLIR imagery

7

author
P. Bijl

date
17 September 1996

number of pages : 47 (incl. appendices, excl. distribution list)

TNO Human Factors Research Institute
is part of TNO Defence research
to which also belong:
TNO Physics and Electronics Laboratory
TNO Prins Maurits Laboratory

Netherlands Organization for
Applied Scientific Research

Managementuittreksel TNO Technische Menskunde, Soesterberg

titel : Acquisition of sea targets. Part 1: Observer performance and "ACQUIRE" model predictions for air-to-surface FLIR imagery (Doelacquisitie op zee. Deel 1: Waarnemingsprestaties en "ACQUIRE" modelvoorspellingen voor IR air-to-surface beelden)
auteur : Dr. P. Bijl
datum : 17 september 1996
opdrachtnr. : A94/KM/315
IWP-nr. : 786.1
rapportnr. : TM-96-A037

Bij de KM bestaat de behoefte om te weten hoe goed menselijke waarnemers met behulp van elektro-optische hulpmiddelen, met name IR waarnemingsapparatuur, in staat zijn om oppervlaktedoelen op zee de classificeren en te identificeren. Dergelijke informatie kan bijvoorbeeld van belang zijn bij het plannen van missies met het Lockheed P-3C ORION patrouillevliegtuig. In het kader hiervan onderzoekt FEL-TNO in opdracht van de KM de mogelijkheden van een operator aid. Een betrouwbaar doelacquisitiemodel zou deel uit moeten gaan maken van een dergelijke operator aid.
Bestaande modellen zijn niet betrouwbaar, zeker niet waar het gaat om doelacquisitie op zee. Deze modellen zijn namelijk ontwikkeld voor landdoelen, en niet beproefd voor zeesituaties. Om deze modellen te toetsen en eventueel aan te passen, of om misschien een geheel nieuw model te ontwikkelen, is het noodzakelijk dat er meetgegevens ter beschikking komen die aangeven hoe goed waarnemers werkelijk in staat zijn om zeedoelen te classificeren en identificeren. Voor het bepalen van deze meetgegevens met behulp van een waarnemingsexperiment werd de hulp van TNO-TM ingeschakeld.
Bij voorkeur wordt hiervoor in een veldtest beeldmateriaal van een aantal doeltypen verzameld, waarbij onder verschillende weerscondities systematisch een aantal belangrijke variabelen, zoals afstand en oriëntatie van het doel, worden gevarieerd. In het laboratorium kan dan voor een groot aantal waarnemers de classificatie- en identificatieprestatie op de beelden worden bepaald. Een dergelijke veldoperatie vergt echter grote inspanning.
Omdat een dergelijke aanpak binnen het huidige project niet haalbaar is, werd besloten om gebruik te maken van bestaand IR-beeldmateriaal, dat de afgelopen jaren door TNO-FEL vanuit de ORION werd verzameld. Aangezien de hoeveelheid bruikbaar beeldmateriaal beperkt is, werd ter aanvulling een experiment uitgevoerd met visuele beelden van zeedoelen die door middel van een simulator werden gegenereerd. Ten opzichte van het eerste experiment heeft dit het nadeel dat het beeldmateriaal niet geheel overeenkomt met de werkelijke situatie op zee, maar het grote voordeel dat systematisch het effect van allerlei belangrijke variabelen op de acquisitieprestatie kon worden onderzocht. Beide experimenten tezamen moeten voldoende inzicht verschaffen in de betrouwbaarheid van doelacquisitiemodellen.
In het huidige rapport wordt het eerste waarnemingsexperiment met het reële IR-beeldmateriaal beschreven. Uit de resultaten werden allereerst enkele vuistregels afgeleid. Als voorbeeld kan worden gesteld dat met de betreffende sensor vanuit de ORION bij mooi weer een S-fregat in zijaanzicht op ongeveer 14 km met 50% kans kan worden geclassificeerd, en op ongeveer 7 km geïdentificeerd. Bij vooraanzicht is dit respectievelijk ongeveer 5 en 2 km. Voor andere doeltypen gelden andere afstanden. De Tydeman, bijna evengroot als het S-fregat, werd in zijaanzicht bijvoorbeeld pas op 7 km door de waarnemers in 50% van de gevallen correct geclassificeerd en op 4 km geïdentificeerd, net als een veel kleinere vissersboot.
De resultaten werden tevens gebruikt om het meest gebruikte doelacquisitiemodel, ACQUIRE, te beproeven. Hiervan bestaan twee versies. De nieuwste versie voldoet op een aantal punten beter dan de oude, met name waar het gaat om het voorspellen van het effect van doeloriëntatie op de waarnemingsprestaties. Het blijkt echter dat dit model zowel de classificatie- als de identificatie-afstanden voor zeedoelen gemiddeld over alle condities met bijna een factor 2 *overschat*. Dit betekent mogelijk dat de in het experiment gebruikte zeedoelen relatief meer op elkaar lijken (b.v. doordat ze minder specifieke details hebben) dan de landdoelen waarvoor het model is geijkt. Ook is gebleken dat het model niet in staat is om allerlei effecten goed te voorspellen. Het gevolg is dan ook, dat de verhouding tussen de gemeten en de door het model voorspelde acquisitie-afstanden per conditie zeer verschilt: deze verhouding loopt uiteen tussen ongeveer 0.16 en 1.6. De conclusie is, dat ACQUIRE beperkt bruikbaar is voor zeedoelen: door een correctiefactor is voor de incorrecte voorspelling van de *gemiddelde* acquisitie-afstand te compenseren; voor het voorspellen van prestaties voor individuele condities is het model onbruikbaar.
Meer gegevens wat betreft doelacquisitie op zee en de toepasbaarheid van doelacquisitiemodellen zullen worden gepresenteerd in deel 2 van deze studie.

Managementuittreksel TNO Technische Menskunde, Soesterberg

titel : Acquisition of sea targets. Part 1: Observer performance and "ACQUIRE" model predictions for air-to-surface FLIR imagery (Doelacquisitie op zee. Deel 1: Waarnemingsprestaties en "ACQUIRE" modelvoorspellingen voor IR air-to-surface beelden)
auteur : Dr. P. Bijl
datum : 17 september 1996
opdrachtnr. : A94/KM/315
IWP-nr. : 786.1
rapportnr. : TM-96-A037

Bij de KM bestaat de behoefte om te weten hoe goed menselijke waarnemers met behulp van elektro-optische hulpmiddelen, met name IR waarnemingsapparatuur, in staat zijn om oppervlaktedoelen op zee de classificeren en te identificeren. Dergelijke informatie kan bijvoorbeeld van belang zijn bij het plannen van missies met het Lockheed P-3C ORION patrouillevliegtuig. In het kader hiervan onderzoekt FEL-TNO in opdracht van de KM de mogelijkheden van een operator aid. Een betrouwbaar doelacquisitiemodel zou deel uit moeten gaan maken van een dergelijke operator aid.
Bestaande modellen zijn niet betrouwbaar, zeker niet waar het gaat om doelacquisitie op zee. Deze modellen zijn namelijk ontwikkeld voor landdoelen, en niet beproefd voor zeesituaties. Om deze modellen te toetsen en eventueel aan te passen, of om misschien een geheel nieuw model te ontwikkelen, is het noodzakelijk dat er meetgegevens ter beschikking komen die aangeven hoe goed waarnemers werkelijk in staat zijn om zeedoelen te classificeren en identificeren. Voor het bepalen van deze meetgegevens met behulp van een waarnemingsexperiment werd de hulp van TNO-TM ingeschakeld.
Bij voorkeur wordt hiervoor in een veldtest beeldmateriaal van een aantal doeltypen verzameld, waarbij onder verschillende weerscondities systematisch een aantal belangrijke variabelen, zoals afstand en oriëntatie van het doel, worden gevarieerd. In het laboratorium kan dan voor een groot aantal waarnemers de classificatie- en identificatieprestatie op de beelden worden bepaald. Een dergelijke veldoperatie vergt echter grote inspanning.
Omdat een dergelijke aanpak binnen het huidige project niet haalbaar is, werd besloten om gebruik te maken van bestaand IR-beeldmateriaal, dat de afgelopen jaren door TNO-FEL vanuit de ORION werd verzameld. Aangezien de hoeveelheid bruikbaar beeldmateriaal beperkt is, werd ter aanvulling een experiment uitgevoerd met visuele beelden van zeedoelen die door middel van een simulator werden gegenereerd. Ten opzichte van het eerste experiment heeft dit het nadeel dat het beeldmateriaal niet geheel overeenkomt met de werkelijke situatie op zee, maar het grote voordeel dat systematisch het effect van allerlei belangrijke variabelen op de acquisitieprestatie kon worden onderzocht. Beide experimenten tezamen moeten voldoende inzicht verschaffen in de betrouwbaarheid van doelacquisitiemodellen.
In het huidige rapport wordt het eerste waarnemingsexperiment met het reële IR-beeldmateriaal beschreven. Uit de resultaten werden allereerst enkele vuistregels afgeleid. Als voorbeeld kan worden gesteld dat met de betreffende sensor vanuit de ORION bij mooi weer een S-fregat in zijaanzicht op ongeveer 14 km met 50% kans kan worden geclassificeerd, en op ongeveer 7 km geïdentificeerd. Bij vooraanzicht is dit respectievelijk ongeveer 5 en 2 km. Voor andere doeltypen gelden andere afstanden. De Tydeman, bijna evengroot als het S-fregat, werd in zijaanzicht bijvoorbeeld pas op 7 km door de waarnemers in 50% van de gevallen correct geclassificeerd en op 4 km geïdentificeerd, net als een veel kleinere vissersboot.
De resultaten werden tevens gebruikt om het meest gebruikte doelacquisitiemodel, ACQUIRE, te beproeven. Hiervan bestaan twee versies. De nieuwste versie voldoet op een aantal punten beter dan de oude, met name waar het gaat om het voorspellen van het effect van doeloriëntatie op de waarnemingsprestaties. Het blijkt echter dat dit model zowel de classificatie- als de identificatie-afstanden voor zeedoelen gemiddeld over alle condities met bijna een factor 2 *overschat*. Dit betekent mogelijk dat de in het experiment gebruikte zeedoelen relatief meer op elkaar lijken (b.v. doordat ze minder specifieke details hebben) dan de landdoelen waarvoor het model is geijkt. Ook is gebleken dat het model niet in staat is om allerlei effecten goed te voorspellen. Het gevolg is dan ook, dat de verhouding tussen de gemeten en de door het model voorspelde acquisitie-afstanden per conditie zeer verschilt: deze verhouding loopt uiteen tussen ongeveer 0.16 en 1.6. De conclusie is, dat ACQUIRE beperkt bruikbaar is voor zeedoelen: door een correctiefactor is voor de incorrecte voorspelling van de *gemiddelde* acquisitie-afstand te compenseren; voor het voorspellen van prestaties voor individuele condities is het model onbruikbaar.
Meer gegevens wat betreft doelacquisitie op zee en de toepasbaarheid van doelacquisitiemodellen zullen worden gepresenteerd in deel 2 van deze studie.

REPORT DOCUMENTATION PAGE

1. DEFENCE REPORT NUMBER (MOD-NL)	2. RECIPIENT'S ACCESSION NUMBER	3. PERFORMING ORGANIZATION REPORT NUMBER
RP 96-0177		TM-96-A037
4. PROJECT/TASK/WORK UNIT NO.	**5. CONTRACT NUMBER**	**6. REPORT DATE**
786.1	A94/KM/315	17 September 1996
7. NUMBER OF PAGES	**8. NUMBER OF REFERENCES**	**9. TYPE OF REPORT AND DATES COVERED**
47	16	Final

10. TITLE AND SUBTITLE

Acquisition of sea targets. Part 1: Observer performance and "ACQUIRE" model predictions for air-to-surface FLIR imagery

11. AUTHOR(S)

P. Bijl

12. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

TNO Human Factors Research Institute
Kampweg 5
3769 DE SOESTERBERG

13. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Director of Navy Research and Development
P.O. Box 20702
2597 PC DEN HAAG

14. SUPPLEMENTARY NOTES

15. ABSTRACT (MAXIMUM 200 WORDS, 1044 BYTE)

In this study, two laboratory experiments were carried out to test how well human observers using an electro-optical (E/O) viewing device, are able to identify or classify sea targets. Such knowledge is of interest to evaluate the applicability of target acquisition (TA) models for a sea environment. Most models are designed and tested for ground targets and backgrounds, and their reliability for acquisition of sea targets is unknown. In the first experiment, which is reported here, observer performance was measured on real thermal infrared (FLIR) imagery that was collected on ORION flights. Experienced observers from several Royal Dutch Navy ORION and LYNX helicopter squadrons participated in this experiment. First, some rules-of-thumb were deduced from the data. For example, with the sensor that was used in the experiment and good atmospheric conditions, an S-frigate in side view may be classified (50% correct) at 14 km and identified at 7 km. For the same target in front view these ranges are about 5 km and 2 km, respectively. For the Tydeman and a Fishing Boat in side view, the 50% correct classification and identification ranges are about 7 and 4 km. Second, the results of the experiment were used to test a widely used TA model, ACQUIRE. Important differences were found between measured and predicted performance. On average, the most recent version of ACQUIRE overestimates acquisition ranges by a factor of 2. Further, the model does not give accurate predictions for individual situations: the ratio between measured and predicted acquisition range depends largely on the circumstances and varies between 0.16 and 1.6 (95%-criterion). The relation between acquisition probability and target range, and the ratio between identification and classification range, are correctly predicted by the model. Apparently, these relations are the same for sea and ground targets. The most recent version of ACQUIRE is preferable to the old version, which is still widely used. The model may be used to predict mean acquisition performance for sea targets, if the predicted ranges are corrected by a factor 0.50. In the second, more extensive experiment, the applicability of models for acquisition of sea targets are tested in more detail. The results will be reported in Part 2 of this study.

16. DESCRIPTORS

Infrared
Model Evaluation
Navy
Observer Performance
Target Acquisition

IDENTIFIERS

Air-to-Surface

17a. SECURITY CLASSIFICATION (OF REPORT)	17b. SECURITY CLASSIFICATION (OF PAGE)	17c. SECURITY CLASSIFICATION (OF ABSTRACT)

18. DISTRIBUTION/AVAILABILITY STATEMENT

Unlimited availability

17d. SECURITY CLASSIFICATION (OF TITLES)

CONTENTS

Report No.: TM-96-A037

Title: Acquisition of sea targets. Part 1: Observer performance and "ACQUIRE" model predictions for air-to-surface FLIR imagery

Author: Dr. P. Bijl

Institute: TNO Human Factors Research Institute
Group: Perception

Date: September 1996

DO Assignment No.: A94/KM/315

No. in Program of Work: 786.1

SUMMARY

In this study, two laboratory experiments were carried out to test how well human observers using an electro-optical (E/O) viewing device, are able to identify or classify sea targets. Such knowledge is of interest to evaluate the applicability of target acquisition (TA) models for a sea environment. Most models are designed and tested for ground targets and backgrounds, and their reliability for acquisition of sea targets is unknown. In the first experiment, which is reported here, observer performance was measured on real thermal infrared (FLIR) imagery that was collected on ORION flights by TNO Physics and Electronics Laboratory (TNO-FEL) in The Hague. Experienced observers from several Royal Dutch Navy ORION and LYNX helicopter squadrons participated in this experiment. First, some rules-of-thumb were deduced from the data. For example, with the sensor that was used in the experiment and good atmospheric conditions, an S-frigate in side view may be classified (50% correct) at 14 km and identified at 7 km. For the same target in front view these ranges are about 5 km and 2 km, respectively. For the Tydeman and a Fishing Boat in side view, the 50% correct classification range is about 7 km and the identification range is 4 km. Second, the results of the experiment were used to test a widely used TA model, ACQUIRE. Important differences were found between measured and predicted performance. On average, the most recent version of ACQUIRE overestimates classification and identification ranges by a factor of 2. Further, the model does not give accurate predictions for individual situations: the ratio between measured and predicted acquisition range depends largely on the circumstances and varies between 0.16 and 1.6 (95%-criterion). The relation between acquisition probability and target range, and the ratio between identification and classification range, are correctly predicted by the model. Apparently, these relations are the same for sea and ground targets. The most recent version of ACQUIRE is preferable to the old version, which is still widely used. The model may be used to predict *mean* acquisition performance for sea targets, if the predicted ranges are corrected by a factor 0.50. In the second, more extensive experiment, the applicability of models for acquisition of sea targets are tested in more detail. The results will be reported in Part 2 of this study.

Rap.nr. TM-96-A037

TNO Technische Menskunde
Soesterberg

Doelacquisitie op zee. Deel 1: Waarnemingsprestaties en "ACQUIRE" modelvoorspellingen voor IR air-to-surface beelden

P. Bijl

SAMENVATTING

In opdracht van de KM werden in het laboratorium twee waarnemingsexperimenten uitgevoerd waarin werd onderzocht hoe goed menselijke waarnemers in staat zijn om met behulp van elektro-optische hulpmiddelen, oppervlaktedoelen op zee te classificeren en identificeren. Dergelijke gegevens zijn gewenst om de bruikbaarheid te onderzoeken van doelacquisitiemodellen voor waarneming op zee. Deze modellen zijn veelal ontwikkeld en getest voor landsituaties, maar over de betrouwbaarheid van de voorspellingen voor zeedoelen is weinig bekend. In het eerste experiment, waarover hier wordt gerapporteerd, werd gebruik gemaakt van bestaand IR-beeldmateriaal, dat de afgelopen jaren door TNO-FEL vanuit het Lockheed P-3C ORION patrouillevliegtuig van de KM werd verzameld. Specialisten uit de bemanning van de ORION en de LYNX helikopter van de KM deden dienst als waarnemers. Uit de resultaten werden allereerst enkele vuistregels afgeleid. Als voorbeeld kan worden gesteld dat met de betreffende sensor vanuit de ORION bij mooi weer een S-fregat in zijaanzicht op ongeveer 14 km met 50% kans kan worden geclassificeerd, en op ongeveer 7 km geïdentificeerd. Bij vooraanzicht is dit respectievelijk ongeveer 5 en 2 km. Voor andere doeltypen gelden andere afstanden. De Tydeman, bijna evengroot als het S-fregat, werd in zijaanzicht bijvoorbeeld pas op 7 km door de waarnemers in 50% van de gevallen correct geclassificeerd en op 4 km geïdentificeerd, net als een veel kleinere vissersboot. Vervolgens werd met de resultaten het meest gebruikte doelacquisitiemodel, ACQUIRE, beproefd. De verschillen tussen de voorspellingen en de gemeten prestaties blijken aanzienlijk. Ten eerste zijn de voorspelde classificatie- en identificatie-afstanden voor zeedoelen, gemiddeld over alle condities, een factor 2 te hoog. Verder blijkt het model niet geschikt voor het voorspellen van prestaties voor individuele situaties: de verhouding tussen de gemeten en de door het model voorspelde acquisitie-afstanden verschilt zeer per situatie en loopt uiteen tussen ongeveer 0,16 en 1,6 (95%-onzekerheidsinterval). De wijze waarop de acquisitiekans afneemt met de afstand tot het doel, en de relatie tussen de identificatie- en classificatie-afstanden, worden door het model wel goed voorspeld. Blijkbaar zijn deze relaties voor zee- en landdoelen hetzelfde. Bovenstaande resultaten gelden voor de meest recente versie van ACQUIRE, die iets beter blijkt te voldoen dan de oude versie. ACQUIRE is geschikt te maken voor voorspelling van de *gemiddelde* acquisitie-afstand door een afstandscorrectiefactor van 0,5 in te voeren. Met behulp van het tweede, veel uitgebreidere experiment zal de toepasbaarheid van modellen voor doelacquisitie op zee nog gedetailleerder worden onderzocht. Hierover wordt gerapporteerd in deel 2 van deze studie.

1 INTRODUCTION

The Royal Dutch Navy has expressed a desire to know how well human observers using an electro-optical (E/O) viewing device, are able to identify or classify sea targets. Such knowledge may be of interest, for example, for mission planning with the Lockheed P-3C ORION patrol aircraft. TNO-FEL investigates the possibilities of an operator aid for the ORION which incorporates a validated target acquisition (TA) model. A TA model predicts the relationship between target range and the probability of correct detection, classification and identification on the basis of the properties of target and background, the atmospheric conditions, and the physical properties of the viewing device used.

There exist a variety of TA models, but their applicability for acquisition of sea targets is unknown. One of the problems is that observer performance data for sea targets, necessary to validate the models, are rare (e.g. Luria *et al.*, 1979). It is expected, however, that most existing models are not directly applicable to sea targets, since they are designed for ground targets and backgrounds.
In the first place, the length-width-height proportions for sea targets can be much more extreme than for ground targets (the proportions for a warship are typically 10:1:1). Hence, an important effect of aspect angle on acquisition performance may be expected. For example, one might expect that acquisition ranges for a ship in side view are longer than for a ship in front view. The most commonly used target acquisition model, however, the 1-dimensional version of the NVESD Static Performance Model (Ratches, 1976; Ratches *et al.*, 1981) predicts equal acquisition ranges for these two situations, since the acquisition performance module takes the minimum target dimension from the observer's point of view as a characteristic size. A recent version of this model (Scott, 1990) takes the square-root area of the projection of the target as characteristic size, which means that it indeed predicts longer ranges for targets in side view than in front view. The square-root area dependency seems a reasonable assumption, but is based on a limited amount of studies (Obert *et al.*, 1990). The acquisition modules for the two versions of the NVESD Model will be called 1-D and 2-D ACQUIRE in the sequel.
A second problem with most models is that they are tuned to performance data for ground-to-ground and air-to-ground target acquisition. Apart from target geometry, other target and background characteristics are also very different for ground and sea targets. Furthermore, the similarity between different targets has a large effect on the acquisition range, and this similarity may be different for the ground targets and sea targets that are of interest. A measure of similarity is not incorporated in the models. This means that, even if the models are in principle applicable to a sea environment, they still have to be calibrated before they may reliably predict acquisition ranges.

The purpose of the present research is twofold. First, we want to determine observer identification and classification performance for sea targets, and second, we want to test existing models. If existing models turn out to be inaccurate for acquisition of sea targets, they may be calibrated or new models may be developed on the basis of the observer performance data.
Two experiments were carried out. In a first experiment, which is described in this report, observer performance was measured on real thermal infrared (FLIR) imagery that was

collected on ORION flights by TNO Physics and Electronics Laboratory (TNO-FEL) in The Hague. Experienced observers from several Royal Dutch Navy ORION and LYNX helicopter squadrons participated in this experiment. The data are used to evaluate two frequently used versions of ACQUIRE for air-to-sea target acquisition. The number of images available for the first experiment was limited, and only some aspects of target acquisition could be tested.
In a second, more extensive, observer experiment, which is described in a separate report (Bijl, 1996) acquisition performance with a *visual* sensor was measured for ship targets that were generated using a simulator. In this experiment, target type, target orientation, depression angle, target contrast and range were varied systematically to be able to quantify the effects of each of these variables on acquisition performance independently.

In chapter 2 of this report, the selection and preparation of the imagery will be discussed. In chapter 3, the ACQUIRE model will be described, as well as the assessment of the input data and the results of the model calculations. The observer performance experiment will be described in chapter 4, and the results of this experiment will be presented in chapter 5. In chapter 6, the ACQUIRE model will be compared with the observer data. The results will be discussed in chapter 7. The reader who is interested in the methods and the results of the observer experiment, but less in the mathematical details of the model and the validation, may consider skipping chapters 3 and 6.

2 IMAGERY

2.1 Image recordings

A large number of air-to-sea recordings of ship targets were collected on ORION flights by TNO Physics and Electronics Laboratory (TNO-FEL) for several studies (see for example De Jong, 1994). The imagery was not specifically recorded for a target acquisition experiment, and only a small fraction of the recordings turned out to be usable for this purpose. All imagery that was used in the present experiment was recorded from a $8-12$ μm thermal imager on MII tape. The Field Of View was 15×9 degrees. A number of target approaches were selected, and a number of short sequences (5 seconds) of each approach were copied to an analogue video disc (see § 4.2) for use in the laboratory experiments.

2.2 Image selection and preparation

2.2.1 Selection criteria

In a target identification and classification experiment, which is carried out to validate a TA model, a number of conditions have to be satisfied:
- the input data to run the target acquisition model (target range, target dimensions, thermal contrast, atmospheric conditions, sensor characterization) have to be known
- the target type in the images should be known (see § 3.4)

- targets should preferably in the 'threshold region', i.e. at ranges where classification and identification are not too easy and not too difficult
- the experimental conditions (target set, sensor type) have to be representative for the conditions under which the model is applied
- the target set should allow a useful definition of the different acquisition levels (identification and classification)
- the image set should be more or less balanced over the targets and other conditions.

For the ORION recordings, not all the data that are required for a model validation are known. First, target ranges and aspect angles were not recorded, and target dimensions are only known accurately for some targets (the ships from the Royal Dutch Navy). In ACQUIRE, it is sufficient to know the target *angular* dimensions (see § 3.3), and these can be estimated from the imagery (§ 3.4.2). Target range and aspect angle (which are convenient to plot the data but are not required for the validation of ACQUIRE) can be calculated if the absolute and angular dimensions of the target are known. If the absolute dimensions are not known, dimensions that are typical for the type of target in the image may be used.
Second, target thermal contrast and the atmospheric transmission are unknown. The MRTD of the sensor is known, but gain and level settings were not recorded. As a result, the signal-to-noise ratio (SNR) of the targets and the corresponding MRTD threshold spatial frequency, which is required to make ACQUIRE predictions, cannot be calculated. Only for high contrast targets, for which the SNR has a negligible effect on the corresponding MRTD spatial frequency, reliable ACQUIRE calculations can be made (see also § 3.3), and for this reason, only *high contrast* targets were used in the experiment.
For some ships, a large amount of imagery was available, including approach runs from several directions towards the target, or circle runs around the target (e.g. for the ships from the Royal Dutch Navy). These runs are very convenient to test a target acquisition model, because a series of image sequences can be selected from distant (identification and classification very difficult) to nearby (acquisition easy). Many other target recordings, however, were only made at close range (too easy to be used in the experiment) or long range (too difficult). Furthermore, often the target types were not documented, and had to be identified from the image. For these targets, only image sequences could be selected from a recording if at any moment during the approach the target could be identified by the experimenter with a 100% certainty.
For the observer experiment, it is further important to minimize the possibility of *picture recognition*. Picture recognition occurs if many sequences from the same target approach are presented, or if there is any relation between target type and, for example, aspect angle, weather type, sea state, sensor type or sensor settings. It is also important that the occurrence of the different targets in the experiment is more or less balanced. As a result, only a limited set of images could be used in the observer experiment.

2.2.2 Image set

A target set of six different ship types was chosen. A set of 137 image sequences from 22 runs were selected on the basis of the criteria mentioned in § 2.2.1. A second set of 35 sequences were selected for a short observer training. The target types are listed in Table I.

Table I The six targets that are used in the experiment. See text for further details.

target type	class
S-frigate	frigate, Dutch Navy
M-frigate	frigate, Dutch Navy
Fishing boat	small vessel, civilian
Coaster	small vessel, civilian
Tydeman	research ship, Dutch Navy
Tanker	large ship, civilian

The ships are divided into three classes: Dutch frigates (warships), small civilian vessels, and other ships. The target set allows a practical definition of different acquisition levels. Target *identification* is defined as naming the correct ship type, such as "S-frigate", or "Tydeman". Target *classification* is defined as naming the correct class: "warship", "small civilian ship" or "other ship".
A short description of the ship types is given below:

Dutch frigates

The Standard frigate (S-frigate) and the Multipurpose frigate (M-frigate) are two types of warships of the Royal Dutch Navy. The two types are quite similar in shape and size, which makes identification to a difficult task. The dimensions of these ships are presented in the first two rows of Table II.
On further consideration, the images of the M-frigates were not used in the experiment, because all M-frigate recordings were made with a different sensor type than the other recordings, and this would have introduced picture recognition (see § 2.2.1). The observers, however, were not informed and the M-frigate was maintained as a target category in the experiment. This was done to allow a comparison of the identification and classification score for the S-frigate with the experiment with simulated targets (Part 2 of this study).

Small civilian vessels

Imagery of several types of Fishing Boats were presented. Most of them were Dutch cutters. The exact dimensions of these boats are unknown. In Table II, typical dimensions of fishing boats are given (in Italic numbers).
Both the classical and the modern type of Coaster were presented. The classical type may easily be confused with a fishing boat. The modern type is quite similar in shape to a tanker, except that it is much smaller. Typical dimensions are given in Table II.

Other ships

The Tydeman is one of the Dutch Navy research ships. The dimensions for this target are given in Table II.

The Tanker is the largest ship type in the set, much larger than all the others, see Table II. In the ORION, observers receive distance information from the radar, which facilitates the identification task for a tanker. In the experiment, no distance information is given. Without distance information a tanker is hard to distinguish from a modern type of coaster.

Table II Dimensions of the six targets (rounded in meters). *Italic numbers* represent typical values for targets of which the exact dimensions are unknown. The target dimensions are only used to make estimates of target ranges and aspect angles. They do *not* affect the comparison between observer scores and ACQUIRE model calculations (see § 3.3).

target type	length (m)	width (m)	funnel height (m)	max. height (m)
S-frigate	131	15	16	32
M-frigate	122	14	15	36
Fishing boat	*40*	*9*	*7*	*14*
Coaster	*90*	*14*	*9*	*16*
Tydeman	90	15	16	26
Tanker	*350*	*60*	*30*	*35*

3 ACQUIRE PREDICTIONS

For the evaluation of a target acquisition model, observer data are only meaningful if the input data for the model, such as the MRTD or the target angular dimensions, can be determined with sufficient accuracy. Therefore, the ACQUIRE calculations for the imagery are made *before* the observer experiment is carried out.

3.1 TA module ACQUIRE

ACQUIRE is the Target Acquisition Module of the NVESD Static Performance Model (Ratches, 1976) and its upgrade called C2NVEO Thermal Imaging Systems Performance Model (Scott, 1990).
ACQUIRE predicts the relation between the range of a target, and the ability of an observer viewing through a thermal system to detect, classify[1] or identify the target. Input variables are: target effective dimension (for a definition, see § 3.2), target thermal contrast, atmospheric transmission, and the MRTD of the viewing system. The MRTD (Minimum Resolvable Temperature Difference) is a threshold performance curve that gives the thermal contrast required by an observer viewing through the device to resolve a 4-bar pattern as a function of spatial frequency.

[1] In ACQUIRE, usually the term 'recognition' is used. The Navy uses the term 'classification' for the same task. In this paper, only 'classification' will be used.

The calculations of the ACQUIRE module are based on the well known 'Johnson Criteria' (Johnson, 1958; Ratches *et al.*, 1981). These criteria link target acquisition performance with the ability to resolve dark bars of a certain spatial frequency and contrast against a uniform background. For example, the model predicts a *classification* probability of 50% if a target is at such a range that an observer is just able to resolve four line pairs over the effective dimension of the target. The higher the resolution of the viewing system, or the larger the target, the longer the range at which four line pairs can be resolved. For a 50% *detection* probability, a resolution of 1 line pair across the effective dimension of the target is required, for *identification* this is 6.4 line pairs. Criteria exist for different levels of probability. The relationships between the number of resolvable line pairs and probability of several acquisition levels are called Target Transfer Probability Functions (TTPF's) (Ratches, 1976). These functions have been established experimentally by averaging over many targets and target orientations. Ratches also indicates the accuracy of the criteria: for a 50% classification probability, the four line pair criterion, mentioned above, would be conservative, whereas a three line pair criterion would be optimistic.

3.2 1-D and 2-D ACQUIRE

In the old, or one-dimensional version of the model, 1-D ACQUIRE, target effective dimension is defined as the minimum dimension of the target from the observer's point of view. The horizontal MRTD (the resolution threshold curve for a *vertical* bar pattern) is used to characterize the performance of the viewing system.

The newer two-dimensional version, 2-D ACQUIRE, takes the square-root area of the projection of the target as effective dimension. A two-dimensional MRTD is introduced to characterize the performance of the viewing system. In the two-dimensional MRTD, the effective 2-D spatial frequency (f_{eff}) is defined as the geometric mean of the horizontal (f_x) and vertical (f_y) MRTD frequency (Scott, 1990):

$$f_{eff} = (f_x \, f_y)^{1/2} \qquad \textbf{(1)}$$

One of the advantages of the 2-D model over the 1-D version is that target area is defined unambiguously, whereas minimum dimension is not. For example, for a ship in side view, the minimum dimension is target height. This gives a considerable degree of freedom: target height can either be defined as bow height, bridge height, or even the maximum height including a mast. In this study, we will take bridge height (or funnel height) as effective dimension for a ship in side view.

3.3 Model simplification

3.3.1 Procedure

For the ORION recordings, thermal contrast and atmospheric reduction are unknown (§ 2.2.1). If apparent thermal contrast is high, the corresponding MRTD threshold spatial

frequency is highly independent of thermal contrast, which means that it is not necessary to know the exact value of this variable to make model predictions. Therefore, only targets were selected for which inherent and apparent thermal contrast with their background is high. It can be shown that for this condition the predicted acquisition probability depends on MRTD cut-off frequency and the ratio between target effective dimension and target range only (Bijl & Valeton, 1994). This relation can be simplified further by expressing the effective target dimension in angular size: target effective dimension D (in mrad) is equal to the ratio between target effective dimension (in m) and target range (in km). The result is a *single* s-shaped function of percentage correct as a function of target effective dimension for each acquisition level, which can be described very well with a curve which is known as the Weibull function. For the prediction of target classification with 1-D ACQUIRE, the relationship is given by:

$$P_{classification,1\text{-}D} = \left(1-2^{-[f_{HMRTD}\cdot D_{MIN}/4]^s} \right) \cdot 100\% \qquad (2)$$

or, inversely:

$$D_{MIN} = \frac{4}{f_{HMRTD}} \cdot \left[-^2\log\left(1-\frac{P_{classification}}{100}\right) \right]^{1/s} \qquad (3)$$

where $P_{classification,\ 1\text{-}D}$ is the predicted probability of a correct classification (in % between 10 and 90), f_{HMRTD} is the cut-off frequency of the horizontal MRTD (in cy/mrad), and D_{MIN} is the minimum target dimension (in mrad). The parameter s determines the steepness of the curve and is set to s=2.06 for an optimal fit to the cycle criteria [given by the TTPF, see Table 1 in Scott (1990)]. It can easily be verified that if D_{MIN} equals 4 resolvable cycles, a classification probability of 50% is predicted.

For identification, the relationship is given by:

$$P_{identification,1\text{-}D} = \left(1-2^{-[f_{HMRTD}\cdot D_{MIN}/6.4]^s} \right) \cdot 100\% \qquad (4)$$

Similarly, equations for acquisition performance predictions with 2-D ACQUIRE can be derived. According to Scott (1990), the cycle criteria for the 2-D model are 25% lower than for 1-D ACQUIRE. This results in the following equations:

$$P_{classification,2\text{-}D} = \left(1-2^{-[f_{2DMRTD}\cdot D_{AREA}/3]^s} \right) \cdot 100\% \qquad (5)$$

and

$$P_{identification,2\text{-}D} = \left(1-2^{-[f_{2DMRTD}\cdot D_{AREA}/4.8]^s} \right) \cdot 100\% \qquad (6)$$

where $P_{classification,\ 2\text{-}D}$ and $P_{identification,\ 2\text{-}D}$ are the predicted probability of a correct classification and identification with 2-D ACQUIRE respectively, f_{2DMRTD} is the cut-off frequency of the two-dimensional MRTD (in cy/mrad), and $D_{AREA} = area^{1/2}$ is the effective target dimension (in mrad). The value of the parameter s is the same for equations 2−6.

3.3.2 Conclusion

In conclusion, only the MRTD cut-off frequency and target effective angular dimension are required to make ACQUIRE predictions of identification and classification probability. In this study, 1-D ACQUIRE predictions are made with the horizontal MRTD cut-off and the target minimum dimension, for 2-D ACQUIRE the 2-D MRTD and the square-root area are used.
All images in the experiment were recorded with the same sensor, which means that the MRTD cut-off frequency is the same for each image. Thus, for the entire image set, predicted acquisition probability vs. angular effective dimension is a single curve (described very well with a simple Weibull function), independent of target range, target dimensions or target orientation. This is especially convenient since for most targets the exact dimensions and their distance to the observer are unknown, and target dimensions are different for each target. The effective dimension in angular size, however, can be estimated from the image. Therefore, predicted and measured acquisition probability are often plotted as a function of (target angular effective dimension)$^{-1}$, which is proportional to target range.

3.4 ACQUIRE input

3.4.1 MRTD

The MRTD of the sensor was obtained from TNO-FEL (De Jong *et al.*, 1991). Small corrections to the cut-off frequency had to be made because the imagery was recorded on MII tape (bandwidth 5.5 MHz) and copied to videodisc (bandwidth 7.5 Mhz). Estimates of the cut-off frequencies are:

horizontal MRTD cut-off frequency:	$f_{HMRTD} = 1.30 +/- 0.15$ cy/mrad
vertical MRTD cut-off frequency:	$f_{HMRTD} = 1.03 +/- 0.15$ cy/mrad
two-dimensional MRTD cut-off frequency:	$f_{2DMRTD} = 1.16 +/- 0.15$ cy/mrad

This MRTD was measured for a stabilized sensor. The sensor platform in the ORION was not stabilized, which means that the sensor performance could be lower in this situation. Therefore, the MRTD was also estimated by measuring the linewidth of thin lines (cables and antennas, for example) in a number of images. The reciprocal of these values give a lower limit of the MRTD cut-off frequency. The results show, that the horizontal cut-off frequency is at least 1.0 cy/mrad, which means that the sensor performance is not influenced drastically by the vibrations of the platform.

The error in the frequency is about 10%. An error in the estimate of the MRTD cut-off affects the comparison of ACQUIRE predictions with mean observer scores. However, the unexplained variance (see § 6.3) is not influenced by an error in the MRTD.

3.4.2 Target effective angular dimensions

As explained in § 3.3, it is sufficient to know the effective angular dimension of the target to make ACQUIRE predictions of classification and identification performance. These dimensions had to be estimated from the images. For each of the 140 image sequences, four target angular dimensions were estimated: minimum dimension, maximum dimension, area and the dimension of a characteristic feature. The maximum dimension was determined to estimate the aspect angle of the target (see § 3.4.3). The dimension of a characteristic feature is measured to improve the precision of the estimates of the effective target dimensions. A characteristic feature is a feature which remains invariant during (at least part of) a target approach, for example target length for a ship that is in side view, or funnel height for a target that rotates during the approach. If the target is approached with a constant speed, a linear or nearly linear relationship is expected between time and the reciprocal of the characteristic dimension, which is proportional to target distance. This was confirmed for most of the runs. By fitting a curve through the data, the time-distance relation could be estimated with more precision, which led to better estimates of the target effective dimensions, especially at long ranges. In a number of cases, the curve had to be extrapolated. The result is that the effective dimensions were determined with an accuracy between 1% and 10%. This accuracy is sufficient for a validation of ACQUIRE, as will be shown later.

3.4.3 Target distance and aspect angle

The target dimensions, given in Table II, were used to estimate target range and aspect angle. These quantities do not affect the validation results of ACQUIRE, but it is convenient to plot acquisition probability as a function of target range and to have an impression of the orientation of the targets. Furthermore, it is interesting to investigate whether the 2-D version of ACQUIRE better predicts the effects of orientation than 1-D ACQUIRE does. Therefore, an estimate of the target aspect angle is required.
Target range is the ratio between the actual target dimension and the angular dimension which was determined in § 3.4.2. Aspect angle was estimated by treating the ships as cuboids and using the ratio between minimum and maximum angular dimension. Only rotations in the horizontal plane were regarded; the depression angle was assumed to be zero (at those ranges where target classification and identification are a difficult task, air-to-sea view is nearly horizontal). Only a rough estimate of the aspect angles was possible, and for the validation of ACQUIRE (§ 6.4) the angles were divided into three classes. These classes are defined in Table III.

Table III Definition of orientation classes. No distinction is made between larboard or starboard. The rightmost column shows that the total number of images is reasonably balanced over the three classes.

orientation class	aspect angles (deg)	number of images
front/rear view	0−15 or 165−180	47
oblique	15−40 or 140−165	40
side view	40−140	50

3.5 Results of the ACQUIRE predictions

Using the estimates of target effective dimensions, target ranges and aspect angles, 1-D and 2-D ACQUIRE predictions were made for all 22 runs using equations (2), (4), (5), and (6) in § 2.3. The results for three runs are presented in Figs 1−3, the complete set is given in Figs A1−A22 in Appendix I. A-figures (left-hand side) represents the identification predictions, B-figures (right-hand side) the classification predictions. In each plot, the target type is given, followed by a number (which is a composition of the original FEL tape-number and a sequence number). Open circles indicate the 1-D ACQUIRE calculations, crosses the 2-D predictions. In the Appendix, the observer data (see chapter 5) are plotted together with the predictions as filled circles. In most runs, target range was the main variable, although the aspect angle usually varied as well. In that case, acquisition probability (in %) is plotted as a function of target range (in km), and the aspect angle of the ship, or a range of aspect angles, is indicated. In the case of a circle run (e.g. Figs A19 and A20) probability is plotted as a function of aspect angle (in degrees), and target range is indicated.

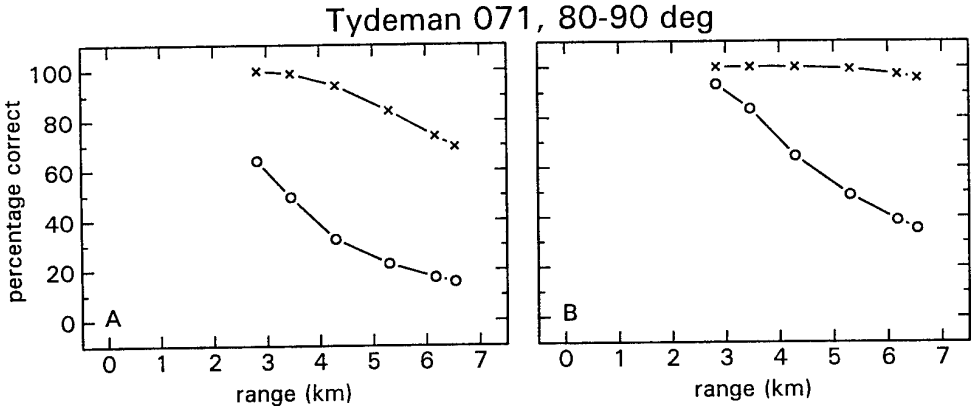


Fig. 1 ACQUIRE probability vs. range predictions for the Tydeman in side view. A: Identification, B: Classification. Open circles: 1-D ACQUIRE, crosses: 2-D ACQUIRE. For a target in side view, 2-D ACQUIRE predicts much longer ranges than the 1-D version does.

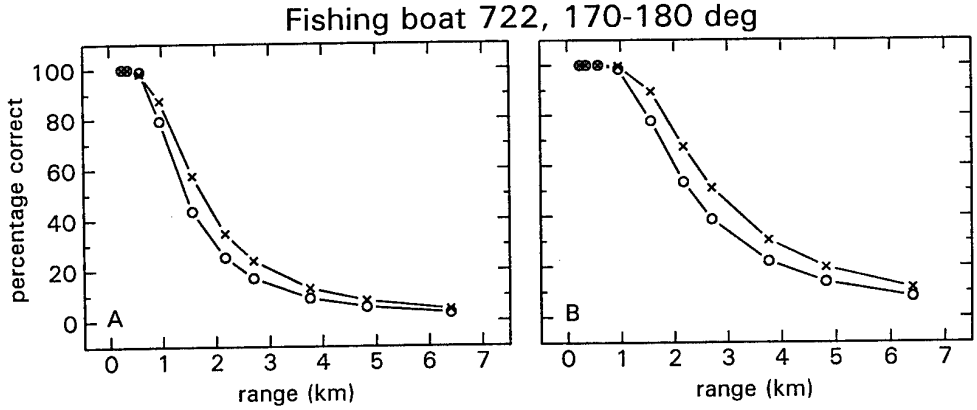


Fig. 2 ACQUIRE probability vs. range predictions for a Fishing Boat in rear view. Symbols as in Fig. 1. If a target is in front or rear view, the two models predict approximately equal acquisition performance.

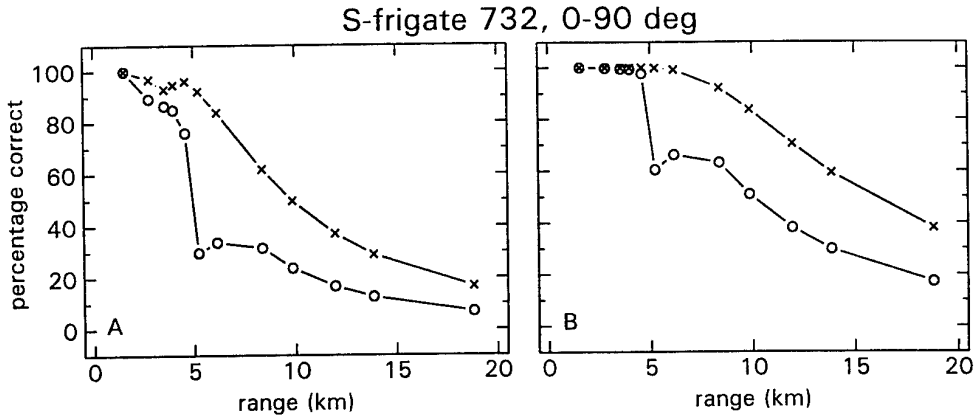


Fig. 3 ACQUIRE probability vs. range predictions for an S-frigate. Symbols as in Fig. 1. Aspect angle varies during the approach of the target. See text for details.

Fig. 1 shows the results for the Tydeman in side view. In Fig. 2, model calculations are given for a Fishing Boat in rear view. Fig. 3 shows the results for an S-frigate of which the aspect angle varies largely during the approach.

The plots show, that 2-D ACQUIRE always predicts a higher acquisition performance than 1-D ACQUIRE. There are three factors that cause differences in predicted range. First, there is a difference of 25% in cycle criteria for the two models (see § 3.3), which results in longer acquisition ranges for the 2-D version. Second, there is a difference between the horizontal and 2-D MRTD. For the sensor used, this difference is very small (see § 3.4.1) and does not have an important effect on the predictions. Third, there is a difference in the definition of the target effective dimension. In most cases, square-root area will be larger than target minimum dimension, which also results in higher performance predictions for the 2-D model. The differences are largest for targets in side view, as is shown in Fig. 1, and smallest for targets in front or rear view, as in Fig. 2.

For most runs, the models predict a gradual monotonous decrease of acquisition probability as a function of target range. However, different curves may be found if the aspect angle of the target varies with target distance. This is seen most clearly in Fig. 3. The predictions of 1-D ACQUIRE show a strong dip or discontinuity at 5 km. The reason is, that the orientation of the target varied during the approach. For most target orientations, target height is the minimum or effective dimension, and the effect of a rotation is rather small. At 5 km distance, however, the target is in front view, which means that target width is the target minimum dimension for a short time. The 2-D ACQUIRE predictions decrease more gradually with target range.

3.6 Conclusions

Errors in the values of the ACQUIRE input parameters (MRTD cut-off frequency and target effective dimension) directly affect the range predictions by the model, and consequently the comparison between model predictions and observer performance. The MRTD cut-off frequencies are known with a precision of about 10%, and the accuracy of target effective

dimension is better than 10%. As a result, the error in the range predictions due to errors in the input variables is less than 20%, which is considered sufficiently small for a meaningful validation study.
Large differences in performance are predicted for different images, and also the predictions made by the two models differ considerably. These factors improve the sensitivity of the validation test.

4 OBSERVER EXPERIMENT

4.1 Experimental setup

A flexible setup was developed in our laboratory to present dynamic video imagery to a number of observers in parallel (see Valeton & Bijl, 1992, 1994). The setup was used to carry out both the experiment with real FLIR imagery and the experiment with visual, simulated imagery which will be discussed in Part 2 of this study. Experienced observers (see § 4.4) participated in both experiments, and were tested in groups of maximally two persons. Civilian observers only participated in the experiment with simulated imagery.
The most important properties of the setup are described below.
The heart of the setup is an analogue video disc system (Sony LVR-6000/LVS-6000P) that was used to present stimuli to the observers. This system is ideally suited for these kind of observation experiments for two reasons. First, it allows the presentation of stimuli to the observers in random order at fast pace. Second, it allows the use of real video sequences as stimuli, which comes as close as possible to real field operation because the image dynamics (spatio-temporal noise and image jitter) are retained. Both stationary and moving targets, recorded from both stationary and moving sensor systems can be displayed realistically. This feature is especially useful for the experiment with real FLIR imagery.
The stimuli were displayed on Sony PVM 122 CE 12 inch monitors (white B4 phosphor) and the contrast and brightness controls were set for optimal linear contrast range before each session. The observers were not allowed to touch the controls.
In the experiment with simulated imagery, Tandy Model 100 notebook computers were used as response panels, and the data were recorded automatically. In the present experiment with real imagery, the observers had to write down their choices on a response sheet after each target presentation. If the experimenter was sure that the responses were given, he pressed a button for the next presentation. The most important reason to collect the data for the two experiments in a different way, is that the target sets are (partly) different. Using the same response panel in both experiments would possibly introduce mistakes or type mismatches by the observers. The experiments were controlled by a PC.
The observers were placed in a dimly lit room and the response panel display or response sheet was illuminated with a small light source. Care was taken that no stray light fell on the monitor screen. The observers were allowed to choose their own optimal viewing distance and to scrutinize the display if they wished to do so.

4.2 Experimental conditions

Experimental sessions with real and simulated imagery were carried out alternately. The complete set of 137 FLIR images were presented twice to the observers in two separate sessions. The first session was preceded by a short training to make the observers accustomed to the experiment.
Each image sequence was presented for 5 seconds. The presentation order was random. No feedback was given, except during the training session.

4.3 Observer task

After each image presentation, the observers were forced to name the target, even if they were not sure which ship was presented. Such a forced-choice procedure has the advantage that observer performance is not biased by observer confidence. They were also asked to indicate whether they were able to identify (I), classify (C) or only detect (D) the target. With the second answer, observer performance is obtained for unforced identification and classification reports, and this procedure is more similar to the target acquisition task in a practical situation (Valeton & Bijl, 1994). With these two responses, four different scores are obtained for each image:

- correct identification (forced): the target type is named correctly
- correct classification (forced): the chosen target belongs to the correct class
- correct identification (unforced): the target type is named correctly, and an (I) was given
- correct classification (unforced): the chosen target belongs to the correct class, and an (I) or (C) was given.

4.4 Statistical error in the observer scores

The fraction of correct identification or classification responses, averaged over the observers and the two sessions, is calculated for each image. If the responses are independent, this leads to a binomial distribution with mean value p (the average fraction of correct responses) and standard deviation of the mean (see also Valeton & Bijl, 1994):

$$\sigma_p = \sqrt{\frac{p(1-p)}{N}} \qquad (7)$$

where N is the total number of presentations. The standard error is maximal if $p=0.5$, and decreases to 0 if p approaches 0 or 1. For example, if $N=14$ (7 observers participated in the experiment, and each image was presented twice), and $p=0.5$, then $\sigma_P=0.13$. If $p=0.2$ or $p=0.8$, $\sigma_P=0.11$. If $p=0$ or $p=1$, $\sigma_P=0$.

4.5 Observers

Seven experienced IR observers from the ORION squadrons VSQ 320 and VSQ 321 (Marine Vliegkamp Valkenburg, The Netherlands), and from a LYNX helicopter squadron (Marine Luchtvaartdienst, Den Helder, The Netherlands), participated in the experiment.
The observers were tested for near vision acuity before entering into the experiment, using the visual acuity chart designed by Walraven *et al.* (1995). For all observers, visual acuity at 60 cm distance (approximately the distance to the CRT on which the images were displayed) was better than 1.3 arcmin^{-1}. This means that the resolution is limited by the sensor system rather than by the visual acuity of the observers.

5 RESULTS OF THE OBSERVER EXPERIMENT

5.1 Data pre-analysis

As was mentioned in § 4.2, each image was presented to the observers two times in two separate sessions. Although no feedback was given, it might be possible that the presentation of a target at different ranges during the first session introduced learning effects. Therefore, it should be tested whether there are differences in scores between the first and the second session. If there are important differences, only the data from the first session may be used.
In a loglinear analysis of variance, the effects of session and image on the forced and unforced identification and classification scores of the observers was determined. Only for the unforced identification scores a significant effect of session ($p<0.05$) was found. This effect, however, was very small (0.5%). The effect of image on the score was 87.5%. Therefore, we may use the data from both sessions.

5.2 Complete set of observer performance data

For all images, the correct scores, averaged over seven observers and two sessions were calculated. Only the unforced identification and classification scores (see § 4.3) will be shown here, since these correspond best to the target acquisition task in a practical situation, and these are in principle the scores that a target acquisition model such as ACQUIRE predicts. However, the results of the analysis of the forced-choice data are quite similar.
The results for two example runs are given in Figs 4 and 5. A-figures represent the identification scores, B-figures the classification results. Fig. 4 shows the observer scores as a function of target range for a Fishing Boat in rear view, and Fig. 5 for an S-frigate of which the aspect angle varies during the approach.
The complete set of observer performance data are plotted together with the ACQUIRE predictions in Figs A1−A22 in Appendix I (Figs 4 and 5 correspond to Figs A9 and A3, respectively). The observer data are plotted as filled circles. An explanation of the figures in given in § 3.5. The statistical error in the data is not plotted, but is approximately 0.11−0.13 for correct scores between 0.20 and 0.80, and decreases to zero for higher or lower scores (see § 4.4).

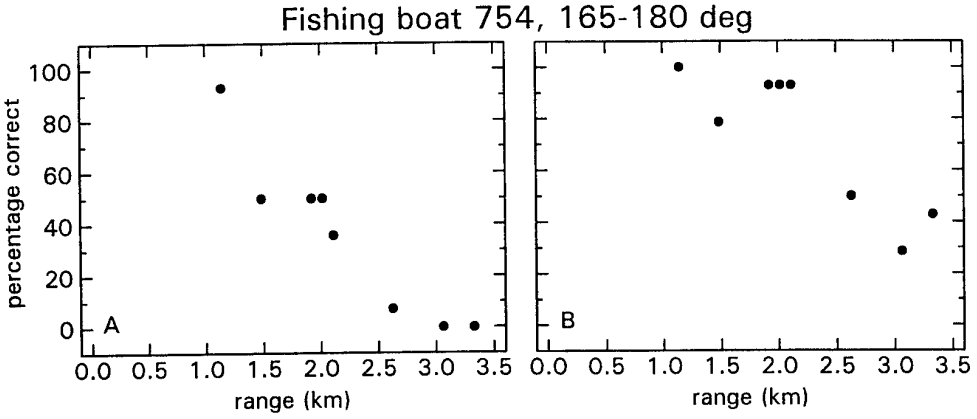


Fig. 4 Observer scores as a function of target range for a Fishing Boat in rear view. A: Identification, B: Classification. If target aspect angle does not vary too much during a target approach, the fraction correct gradually decreases with target range.

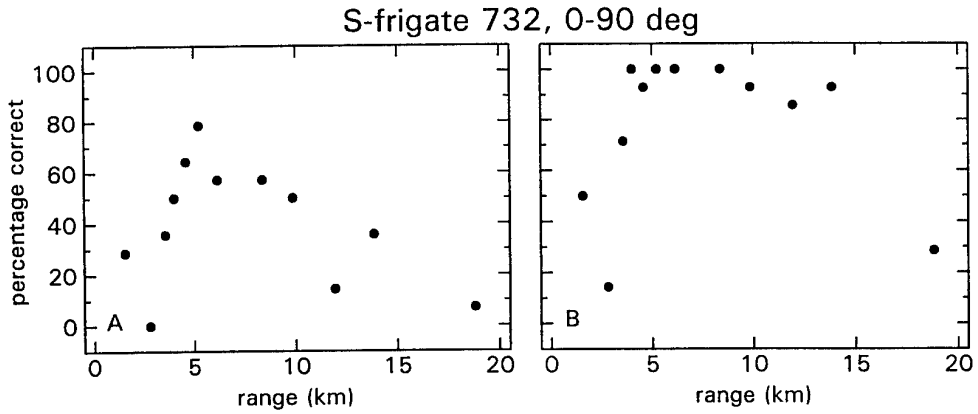


Fig. 5 Observer scores as a function of target range for an S-frigate. A: Identification, B: Classification. Aspect angle varies during the approach of the target, and this has a large effect on acquisition probability. See text for details.

5.3 Qualitative inspection of the observer performance data

The results in Figs A1−A22 show, that there is a large variation in performance between the different images. The acquisition tasks are not too easy (correct score often near 1), or too difficult (acquisition score often near 0). This means that the data set is suitable for a validation of the acquisition models.

If target aspect angle does not vary too much during a target approach, the fraction correct identification and classification gradually decreases with target range. This is shown e.g. in Fig. 4, and in Figs A1, A5, A6, A7, A9, A10, A11, A14, A15, A16, A18, and A21. Most of the variation in score between adjacent data points can be ascribed to the statistical error in the observer scores.

The effect or aspect angle can be quite large. This is for example shown in Fig. 5. In this run, the S-frigate is in side-view (90 degrees) for ranges above 5 km, and between 0 and 20 degrees for ranges below 4 km. Approaching from 18 km to 5 km, identification (Fig. 4a)

and classification performance (Fig. 4b) gradually increase, as was found in the other runs where orientation angle was approximately constant. Approaching from 5 km to 1km, the aspect angle reduces to 0 degrees (front view), and at the same time acquisition performance drops significantly. Other examples are found in Figs A19, A20, and A22. In Fig. A19, target range is approximately constant. Acquisition performance greatly improves as the aspect angle increases from 30 degrees to 90 degrees. In Fig. A22, Tanker 751 is in side view for the three longest ranges, and in front view for the two other ranges. Acquisition performance is highest for the target in side view.
For different target types, considerable differences in acquisition ranges are found. For the S-frigate in Fig. 5 (which is in side view for ranges above 5 km), the 50% identification range is about 7 km, and the 50% classification range is about 16 km. For the Tydeman in Fig. A14, which is also in side-view, the 50% identification and classification ranges are about 4 km. For the Tanker in Fig. A22, which is in side view for ranges between 8 and 14 km, both 50% identification and classification range are longer than 14 km. There are also differences between different runs of targets of the same type. For example, the Fishing Boat number 722 and 754 in Figs A6 and A9 (Fig. 4) are both in rear view, but the 50% identification ranges differ by about a factor of two.

In conclusion, observer performance seems to decrease gradually with target distance if other variables, such as target aspect angle, remain constant. Acquisition ranges are different for different targets. These findings qualitatively agree with the predictions of a TA model such as 1-D or 2-D ACQUIRE. A strong effect of target orientation on acquisition performance is found. Such an effect is predicted by 2-D ACQUIRE but not by 1-D ACQUIRE. There are also differences in performance for different runs of the same target type in the same orientation, and this is an effect that is not predicted by the two versions of ACQUIRE. These differences may be due to factors such as sea state, or other subtle differences that can have a large effect on acquisition performance but are not incorporated in the model.

5.4 Some quantitative results

Before the acquisition model predictions are compared with the observer results, some rough quantitative results are deduced from the observer data to give a feeling of the ranges at which the targets can be identified or classified. These results might, for example, serve as a rule-of-thumb for patrol flights.

5.4.1 Procedure

All data for each target type are plotted in a single graph. When the observer scores are plotted as a function of target range, as in Figs A1−A22, scattered datapoints are expected due to the effect of orientation on acquisition performance. Instead, they are now plotted as a function of (target area)$^{-1/2}$, where area$^{1/2}$ is expressed in angular size (mrad). This quantity is proportional with target range as long as target dimensions and aspect angle are constant (see § 3.4), but compensates for the effect of changes in orientation in the way that 2-D ACQUIRE does. If the square-root area assumption is correct, a single probability vs (target

area)$^{-1/2}$ curve will be found for each target, independent of orientation. From this curve, a 50% correct identification or classification square-root area (in angular size) can be found. This quantity may be converted to a 50% correct range for targets if the absolute target square-root area is known. For a target in side view, absolute target area will be approximated by the product of target length and funnel height, which are given in Table II. For a target in front view the product of target width and funnel height will be used.

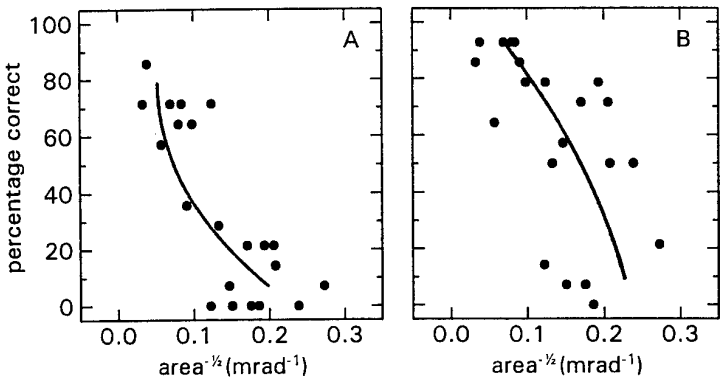


Fig. 6 Entire set of observer scores for the Tydeman as a function of (target area)$^{-1/2}$. A: Identification, B: Classification. A rough overall curve was drawn through the data to estimate the 50% correct (target area)$^{1/2}$. See text for details.

5.4.2 Results

In Fig. 6a and b, all identification and classification scores for the Tydeman are plotted as a function of (target area)$^{-1/2}$. In Figs A23−A27 in Appendix II, similar plots are given for each target type in the set. A rough overall curve was drawn through the data to estimate the 50% correct target square-root area. Estimates of these values are given in Table A1 in Appendix II. Column 2 gives the results for target identification and column 3 for target classification.

The results of the range calculations are given in Table IV. Estimates of the 50% identification range of targets in side and front view are given in columns 2 and 3, respectively. Classification ranges are given in columns 4 and 5. The results indicate that an S-frigate in side view can be classified as a frigate at a distance of about 14 km, and identified at about 7 km. For a fishing boat, these ranges are 7 and 4 km, respectively. Identification of most targets in front view is only possible at a distance below 2 km.

The results in Table IV can serve as rules of thumb. Of course, they are a very rough mean over a number of conditions. For individual situations, the estimates can be far off. Furthermore, these ranges only hold for the sensor that was used in this experiment, and for targets with sufficiently high apparent thermal contrast (i.e. for good atmospheric conditions).

Table IV Rough estimates of 50% identification and classification ranges of the targets in side and front view, based on the data from the observer performance experiment. These estimates can serve as rules of thumb.

target type	50% identification range (km)		50% classification range (km)	
	side view	front view	side view	front view
S-frigate	7	2	14	5
Fishing boat	4	2	7	3
Coaster	4	1.5	5	2
Tydeman	4	1.5	7	3
Tanker	10	4	13	6

6 ACQUIRE VALIDATION

The validation of 1-D and 2-D ACQUIRE will take place at various levels of complexity.
In § 6.1, a qualitative comparison between the observer data and model predictions will be made to give a rough indication of the accuracy of the model predictions.
In § 6.2, the 1-D and 2-D ACQUIRE predictions will be compared to the overall mean score over all runs. Originally, the Johnson criteria were based on mean acquisition performance over a large number of conditions (with ground targets). Therefore, it is useful to check their validity for sea targets.
Finally, a complete quantitative comparison will be made between the set of individual datapoints and the model predictions. As can be seen from the figures in Appendix I, large performance differences are found for different conditions, which may be due, for example, to target type, orientation or probably other factors such as sea state. Only target effective dimension is modeled in ACQUIRE. Thus, the model cannot account for at least part of the variation in the observer data. This part, which is called the "unexplained variance" will be determined in § 6.3. It is of obvious importance to know how large the unexplained variance is. It directly provides a measure of the reliability of the acquisition ranges predicted by the model. If the amount of unexplained variance is small, the model is able to make reliable predictions for individual situations (e.g. the 50% identification range for an M-frigate in side-view from 100 feet above sea level). If the amount of variance is large, the model may only be useful to predict overall mean performance.
In § 6.4, the unexplained variance will be analyzed. Possibly, part of the variance can be ascribed to one or more factors that are not, or incorrectly, modeled in ACQUIRE. If these factors can easily be incorporated, the amount of unexplained variance may be reduced and the model may be improved.

6.1 Qualitative validation of ACQUIRE

In Figs 7−10, the observer performance data and the 1-D and 2-D ACQUIRE predictions are plotted together for four example runs. Fig. 7 shows the data for Fishing Boat 754, Fig. 8 for Tydeman 092, Fig. 9 for S-frigate 732, and Fig. 10 for Coaster 691. The entire data

set is given in Figs A1 – A22 in Appendix I. An explanation of the figures is given in § 3.5 and 5.2.

Comparison of the ACQUIRE calculations with the observer data shows that predictions of both models are sometimes good or reasonable, as in Figs 7 and 8, but also often far off, as in Figs 9 and 10. On the basis of these observations, we conclude that the predictions for individual situations is not very accurate.

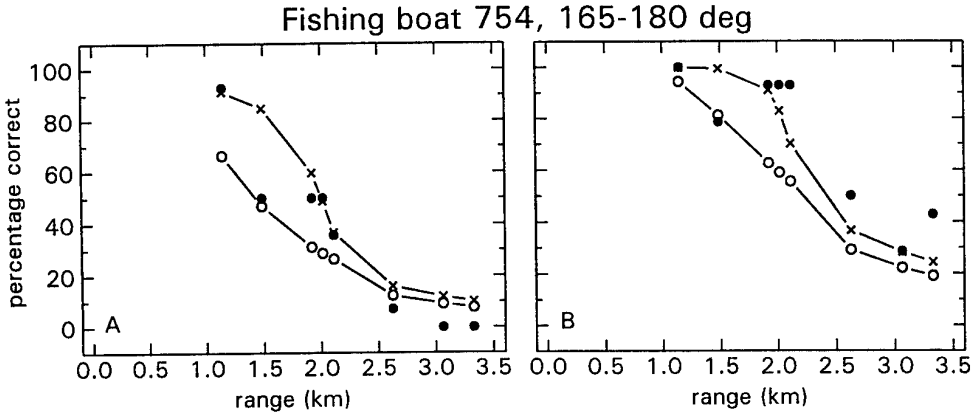


Fig. 7 Observer scores plotted together with the 1-D and 2-D ACQUIRE probability vs. range predictions for Fishing Boat 754. A: Identification, B: Classification. Filled circles: observer data. Open circles: 1-D ACQUIRE. Crosses: 2-D ACQUIRE. For this particular run, the model predictions are good.

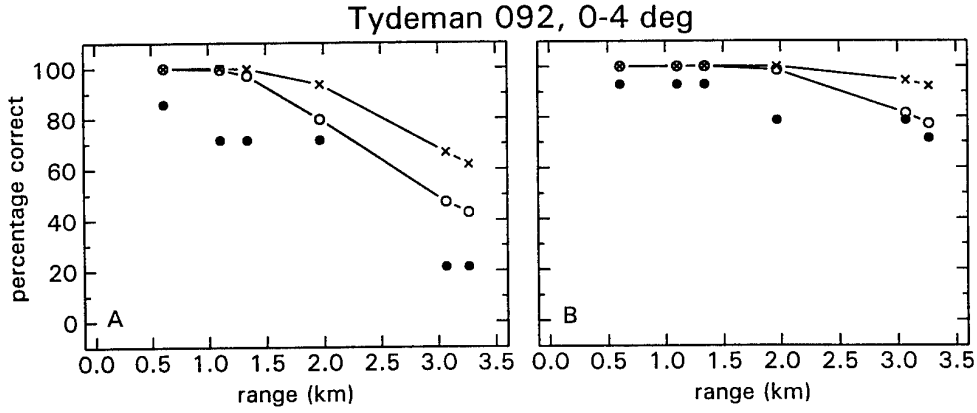


Fig. 8 Observer scores plotted together with the 1-D and 2-D ACQUIRE probability vs. range predictions for Tydeman 092. Symbols as in Fig. 7. For this run, the model predictions are reasonable.

In general, the 2-D model predictions appear to be too optimistic. On average, the predictions by the two models seem reasonable for the Fishing Boat and the S-frigate, but acquisition performance for the Coaster and the Tydeman is overestimated, especially by the 2-D model. Apparently, the effect of target type is significant, although it is not modeled in ACQUIRE.

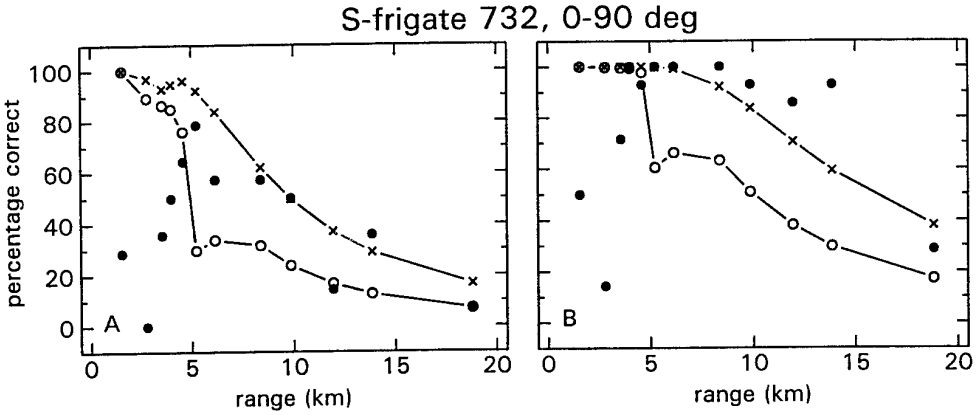


Fig. 9 Observer scores plotted together with the 1-D and 2-D ACQUIRE probability vs. range predictions for S-frigate 732. Symbols as in Fig. 7. The two versions of the model do not compensate sufficiently for the change in target orientation.

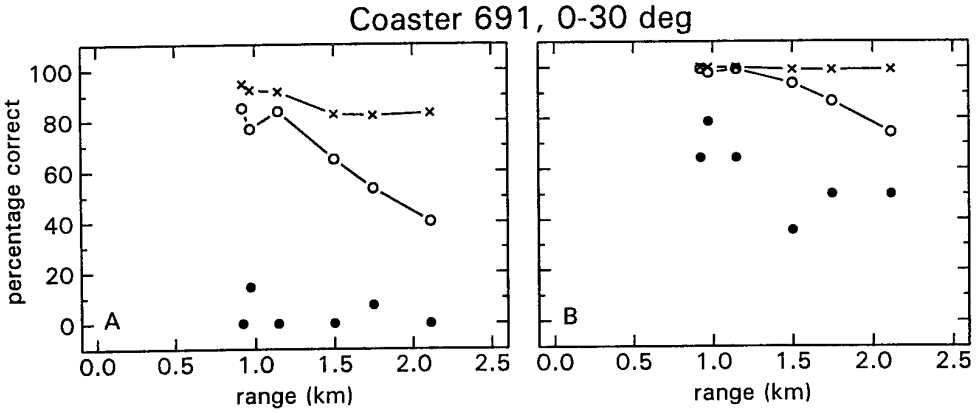


Fig. 10 Observer scores plotted together with the 1-D and 2-D ACQUIRE probability vs. range predictions for Coaster 691. Symbols as in Fig. 7. For this run, the model predictions are far off.

It is not easy to find other systematic effects, although it is clear that the identification and classification score for S-frigate 732 (Fig. 9) at close range are low because the target is in front view (see 5.3), and that the model does not compensate sufficiently for that effect. The same result is found for Tanker 751 in Fig. A22. These effects will be considered in § 6.4.

6.2 Comparison with overall mean observer performance

In this section, a quantitative comparison will be made between the ACQUIRE predictions and mean acquisition performance over a large number of conditions.

As shown in § 3.3, it is convenient to consider the relationship between acquisition probability and (angular effective dimension)$^{-1}$ rather than target range. ACQUIRE predicts a single curve between these two variables for the entire image set, independent of target ranges, dimensions or target orientations.

Acquisition probability will be considered as the independent variable, and (target angular effective dimension)$^{-1}$ as the dependent variable. Mean observer performance will be defined as the relation between acquisition probability, and the geometric average of (target angular effective dimension)$^{-1}$ for all images that correspond to that probability level.
The difference between mean observer performance and the model predictions will be expressed as a *range factor*. This range factor is defined as the ratio between actual acquisition range and predicted acquisition range, which equals the ratio between the predicted effective dimension and the geometric average of the real effective dimensions. For example, if the predicted range is equal to mean acquisition range, the range factor equals 1. If the predictions are optimistic or pessimistic, the range factor is less than 1 or higher than 1, respectively.

The results are plotted in Figs 11 and 12. Figs A show the results for identification, Figs B for classification. In Fig. 11, the observer scores (filled circles) are plotted as a function of the geometric average of (target minimum dimension)$^{-1}$. The 1-D ACQUIRE model predictions are plotted as a solid line. The dashed line indicates an optimal fit of the model to the observer data, using the range factor. In Fig. 12, the observer data are plotted as a function of the geometric average of (target area)$^{-1/2}$, together with the 2-D ACQUIRE predictions (solid line) and an optimal fit for this model (dashed line). The figures show that:

- On average, the 1-D ACQUIRE predictions for identification are reasonable, but the curve for classification is slightly optimistic compared to measured performance. The optimal correction factor is 0.85 for identification. Thus, on average, 1-D ACQUIRE identification range predictions should be multiplied by 0.85 to correspond best to the observer performance data. For classification, the range factor is 0.70.
- 2-D ACQUIRE overestimates mean observer performance. A good fit is obtained for both identification and classification with a range factor of 0.50.
- When corrected by the optimal range factor, the slope of the predicted curves corresponds well to the slope in the data, except maybe for the 1-D ACQUIRE identification curve in Fig. 11a, which seems slightly too shallow. For the lowest probability levels, all predicted curves seem too shallow. The significance of these datapoints, however, is limited (see also the next section).
- Especially for identification, the data is more scattered when plotted as a function of (minimum dimension)$^{-1}$ than when it is plotted as a function of (target area)$^{-1/2}$. This indicates, that target square-root area is a better measure to predict acquisition performance than minimum dimension is. This makes 2-D ACQUIRE a potentially better model for acquisition of sea targets than 1-D ACQUIRE. This finding will be discussed in more detail in § 6.4.

In conclusion, 2-D ACQUIRE overestimates mean observer performance by a factor of 2, 1-D ACQUIRE by a factor of 1.2 (identification) to 1.4 (classification). The range factor is almost equal for the two acquisition levels, which means that the relationship between these levels in the present experiment with sea targets is correctly incorporated in the model (see § 3.1). In general, the slope of the predicted curves corresponds well to the slope in the data, which means that the shape of the TTPF's (§ 3.1), which has been determined for ground targets, is also correct for sea targets.

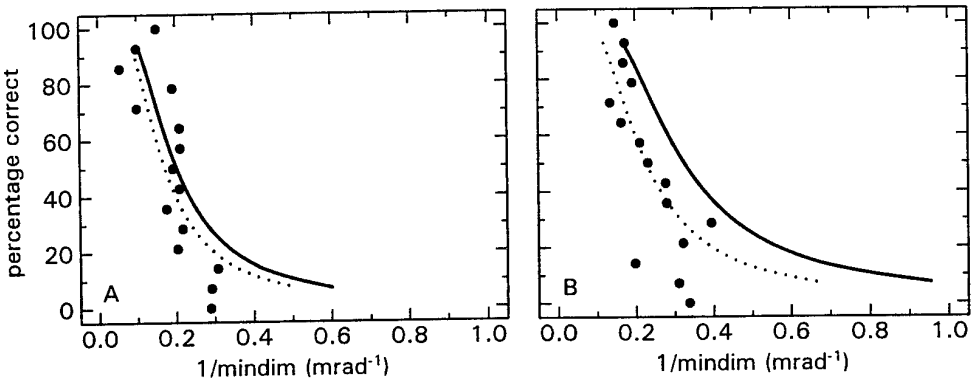


Fig. 11 Comparison of overall mean observer performance and 1-D ACQUIRE predictions. A. Identification. B. Classification. Open circles: Observer score as a function of the geometric average of (target minimum angular dimension)$^{-1}$. Solid line: 1-D ACQUIRE prediction. Dashed line: ACQUIRE predictions, corrected for a range factor. On average, the 1-D ACQUIRE predictions are slightly optimistic. The optimal range factor is 0.85 for identification and 0.70 for classification. See text for details.

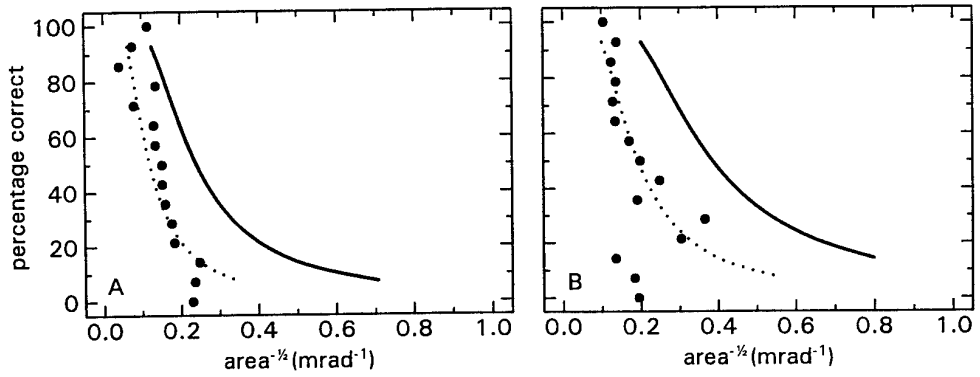


Fig. 12 Comparison of overall mean observer performance and 2-D ACQUIRE predictions. A. Identification. B. Recognition. Open circles: Observer score as a function of the geometric average of (target area)$^{-1/2}$. Solid line: 2-D ACQUIRE prediction. Dashed line: ACQUIRE predictions, corrected for a range factor. 2-D ACQUIRE overestimates mean observer performance, but when the predicted (target area)$^{-1/2}$ is multiplied by a factor of 0.50, the model fits better to the data than the 1-D version. See text for details.

6.3 Comparison with observer performance for individual trials

In the previous section, it was shown that overall mean acquisition performance can be described reasonably well with a monotonously decreasing function of target range or (angular effective dimension)$^{-1}$. The slope in the curves that are predicted by ACQUIRE corresponds well to the data if we correct the predicted ranges or target effective dimensions by a single factor.

In this section, a comparison will be made between the ACQUIRE predictions and acquisition performance for *individual trials*. In § 6.1, it was concluded that these predictions are not very accurate. How (in)accurate the predictions are, will be determined quantitatively to

find the limitations of the model, and to compare the accuracy of 1-D and 2-D ACQUIRE predictions.
Basically, the observer performance data will be regarded as a large set of single probability *vs* range points, and a point by point comparison will be made between actual target and predicted range. This procedure yields a distribution which gives the unexplained variance. This is the variance in the data due to all parameters that were varied in the experiments but which is not predicted by the model. Part of this variance is due to the statistical error in the observer scores. The other part of the unexplained variance provides an indication of the quality of the model predictions for individual trials.

6.3.1 Procedure

The procedure is explained in detail in Bijl and Valeton (1994). Each datapoint represents a probability of correct identification or classification P for a target at range r (or, equivalently, an angular effective dimension D). At this probability level, the model predicts a range r' (or an effective dimension D'). For each datapoint in the set, a ratio r/r' between actual and predicted effective dimension is calculated (this ratio is equal to D'/D). If the model makes a correct prediction, the ratio is $r/r' = 1$. If the predicted range is too short, the ratio will be larger than 1, and if predicted ranges are too large, the ratio will be smaller than 1. It is convenient to transform the values to a log-scale, because the correct predictions are then centred at 0, and an over- or underestimate of target range by the same factor (for example, the predicted range is twice or half the actual range) are represented by equal shifts in opposite directions along the horizontal axis.

A set of datapoints gives a (dimensionless) distribution of log (r/r')-values. Mean and variance of the distribution directly provide a measure of the accuracy of the model. The *mean* of the distribution, $<<\log (r/r')>>$, indicates how well the model predicts overall mean performance. If $<<\log (r/r')>> = 0$, overall mean performance is correctly predicted. A shift of the distribution along the log (r/r')-axis means that, on average, predicted range is too long or too short. In that case, $<<\log (r/r')>>$ provides a range correction factor that will make the model predict overall mean performance correctly. Essentially, this factor should correspond to the range factors that were found in § 6.2.

The *variance* $\sigma^2_{(r/r')}$ in the distribution indicates how well the model predicts observer performance for individual trials. Part of the variance, $\sigma^2_{(r/r'),obs}$, will be due to statistical errors in the observer scores (§ 4.4), because an error in the identification or classification probability P leads to an error in r', and hence in log (r/r'). The error propagation depends on the slope of the probability vs. range function, as is shown in Bijl and Valeton (1994). At very low or high probability levels, where the slope is shallow, a small error in the observer score leads to a large error in predicted range. For this reason only observer scores between 20% and 90% are used in the analysis. It can be shown that for this dataset, the variance in log (r/r') due to the errors in observer scores is approximately 0.005. Another part of the variance can be ascribed to the errors that are made in the estimates of the effective dimensions from the imagery (§ 3.4.1). This error in D (and thus in r) was estimated between 1% and 10%. The variance in the distribution due to this error will be smaller than $[\log(1.10)]^2 = 0.0016$. Thus, the maximum variance due to both the error in observer scores

and in the estimates of the effective dimensions is less than 0.0066 or 0.08^2. This corresponds to a standard deviation in range of about 20%. The remainder of the variance can be ascribed to incorrect predictions by the model.

If the evaluation yields that $\sigma^2_{(r/r')}$ = approx. 0.0066, the model will be accepted because it correctly predicts observer performance (within the experimental error) as a function of the parameters that were varied in the experiment (e.g. target type, target orientation). On the other hand, if $\sigma^2_{(r/r')} >> 0.0066$, the unexplained variance is mainly due to differences between model predictions and actual observer performance. In that case, the 95% confidence interval of the distribution, $[<<\log (r/r')>> - 2\ \sigma_{(r/r')}, <<\log (r/r')>> + 2\ \sigma_{(r/r')}]$, indicates the quality of the model predictions. A wide uncertainty interval indicates that the model is not able to make reliable predictions for individual trials.

6.3.2 Results

In Figs 13 and 14 histograms of the distribution of log (r/r')-values are given for the comparison of the observer performance data and the 1-D and 2-D ACQUIRE predictions, respectively. Figs A show the results for identification, Figs B for classification. Note that the distributions are approximately normal. Mean and variance of the distributions are given in Table V.

It is clear, that the quality of the model predictions of both 1-D and 2-D ACQUIRE for individual trials is not very good. First, the mean of the distribution is below 0, which means that the ranges predicted by the models are too long. This result was already found in § 6.2. The factors that were found in § 6.2 quantitatively agree with the values in Table V. For example, a mean identification shift of -0.07 for identification predictions with 1-D ACQUIRE mean that the model predictions should be corrected by a factor $10^{-0.07}=0.85$. A shift of -0.30 (Classification, 2-D ACQUIRE) means a factor of $10^{-0.30}=0.50$.
Second, the amount of variance is much larger than the variance due to the errors in observer score and in the estimates of the effective target dimensions from the images. For example, a variance of 0.075 means a standard deviation of 0.27 on a log scale, and a confidence interval of $[-0.62, 0.48]$. On a linear scale, this interval is [0.24, 3], spanning a factor of more than 10. In words, there is a 95% probability that the actual identification range for an individual trial falls between 0.24 and 3 times the range that is predicted by 1-D ACQUIRE. The amount of unexplained variance is lower with 2-D ACQUIRE, but even for the identification predictions with this model the interval on a linear scale is [0.19, 1.6], spanning a range of a factor of 8. Thus, even if the model predictions are corrected for the overall mean acquisition range, the actual range may be about 0.3 to 3 times the predicted range.

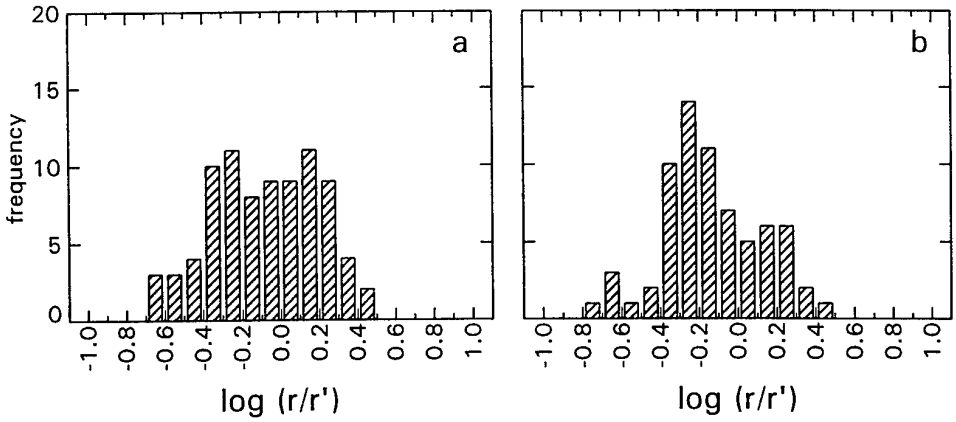


Fig. 13 Histograms of the distribution of log (*r/r'*)-values for the comparison of the observer performance data and the 1-D ACQUIRE predictions. A. Identification. B. Classification. Mean and variance of the distributions are given in Table V. See text for details.

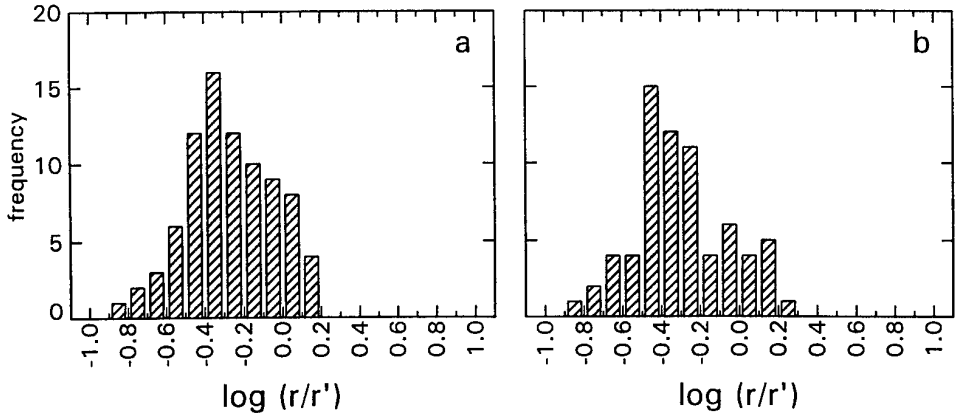


Fig. 14 Histograms of the distribution of log (*r/r'*)-values for the comparison of the observer performance data and the 2-D ACQUIRE predictions. A. Identification. B. Classification. Mean and variance of the distributions are given in Table V. See text for details.

Table V Mean and variance in log (predicted effective dimension/actual effective dimension) for identification and classification with 1-D and 2-D ACQUIRE. The variance due to the statistical error in observer score and target effective dimensions is given in rows 3 and 4, respectively. See text for details.

Model	Identification		Classification	
	Mean	Variance	Mean	Variance
1-D ACQUIRE	−0.07	0.075	−0.14	0.067
2-D ACQUIRE	−0.27	0.054	−0.30	0.064
observer error	–	0.005	–	0.005
effective dimension error	–	0.0016	–	0.0016

In conclusion, the two versions of ACQUIRE are not very good at predicting how observer performance depends on the prevailing conditions. The amount of unexplained variance is lowest for the 2-D version.

6.4 Possible sources of the unexplained variance

In the previous section it was shown that, when the ACQUIRE predictions are compared with actual observer performance for individual trials, there is a large amount of unexplained variance. The variance cannot be ascribed to statistical errors. The conclusion was, that the model does not predict how observer performance really depends on field factors or parameters that were varied in the experiment.
In this section, the effect of several factors on the unexplained variance was determined. One of the main findings is, that the effect of orientation is predicted better by 2-D ACQUIRE than by the 1-D version of the model. This was concluded from the finding that target orientation has a significant effect on the unexplained variance for 1-D ACQUIRE, but not for 2-D ACQUIRE.
A second finding is, that there seems to be no simple modification that would lead to a model that accurately predicts acquisition performance for individual trials. The effect of target type was found to be significant, but remodelling this effect does not greatly improve the model predictions.

6.4.1 Procedure

Using Analysis of Variance, the significance of the effects of target type, target orientation class (see Table III) and probability level on log (r/r') is tested separately. These are the independent variables that are known for the imagery. The analysis was carried out for both identification and classification, and for both 1-D and 2-D ACQUIRE.
If the effect of a variable is *not* significant, it means that the model predicts its effect on acquisition range sufficiently well, and it is not necessary to improve the model for that variable.
If the effect is significant, it means that log (r/r') distributions with different means are found for different values of the independent variable, e.g. for different target types. In that case, the broad distribution of log (r/r')-values that is found for the entire dataset, is in fact a composition of narrower distributions with different means for different values of the variable. If we take the ACQUIRE predictions, and we apply an optimal mean shift (or range correction factor) for each value of the independent variable (i.e. the opposite of the mean that is found), then we have a model which optimally predicts the effect of this variable on acquisition range. For this (new) model, we again estimate the unexplained variance. If this variance is much smaller than the unexplained variance for the original model, it is worthwhile to consider (re-) modelling the effect of the specific variable.

6.4.2 Results

For 1-D ACQUIRE, the effect of target type and orientation class on log (r/r') are significant ($p<0.05$). For identification, also the effect of probability level is significant. For 2-D ACQUIRE, only the effect of target type is significant.
These findings confirm our earlier observation that the effect of orientation is predicted better by the 2-D ACQUIRE than by the 1-D version of the model. Apparently, the model was indeed improved by taking the square-root area instead of the minimum target dimension. In the experiment with simulated imagery, the square-root area assumption can be tested more precisely (see Part 2 of this study). Another finding is that the effect of probability level is not significant, except for the identification range predictions with 1-D ACQUIRE. This agrees with our finding that the slope of the probability versus range function is predicted correctly (§ 6.2, Figs 11 and 12). Only the curve for identification with 1-D ACQUIRE (Fig. 11a) seems shallower than measured.
Both findings indicate that 2-D ACQUIRE is in principle better suited to predict acquisition of sea targets than the 1-D version. This also agrees with the finding that the unexplained variance is lower for the new version (see § 6.3, Table V).
Next, the 2-D ACQUIRE predictions were taken, and optimal range correction factors were applied for each target type separately. For this new model, the amount of unexplained variance is 0.045 for identification (2-D ACQUIRE: 0.054), and 0.038 for classification (0.064). The predictions of this new model for individual situations are better than the original ACQUIRE predictions, but they still are not very accurate: for identification the 95% confidence interval spans a range of a factor 7 (2-D ACQUIRE: 8.5) and for classification a factor of 6 (10).
The remaining amount of unexplained variance has to be ascribed to variables that where not controlled in the experiment (such as sea state). Also interactions, e.g. the combined effect of target type and orientation, may play an important role. The effect of interactions may be determined in the experiment with simulated imagery. Both factors are not modeled in ACQUIRE. In conclusion, there seems to be no simple modification that would lead to a model that accurately predicts acquisition performance for individual trials.

7 DISCUSSION AND CONCLUSIONS

In this study, a set of observer performance data for identification and classification of sea targets was collected, using real FLIR imagery recorded on ORION flights. Further, the data were used to test the applicability of two versions of the most commonly used target acquisition model, ACQUIRE, for acquisition of sea targets.

Observer data

The data show that, if all other variables remain constant, observer performance decreases gradually with target distance. A strong effect of target orientation was found. As was expected, acquisition ranges are different for different targets, but there are also considerable differences for the same target under slightly different conditions. Quantitatively, some

rules-of-thumb were deduced from the data. These are presented in Table IV. For example, with the sensor that was used in the experiment and good atmospheric conditions, an S-frigate in side view may be classified (50% correct) at 14 km and identified at 7 km. For the same target in front view these ranges are about 5 km and 2 km, respectively. For the Tydeman and a Fishing Boat in side view, the 50% correct classification range is about 7 km and the identification range is 4 km.

ACQUIRE model predictions

ACQUIRE qualitatively predicts a number of these findings, but quantitative predictions of the model are not very accurate. On average, measured acquisition ranges are about 0.50 times the ranges that are predicted by the most recent version of the model, 2-D ACQUIRE. Further, the ratio between measured and predicted acquisition ranges are very different for different conditions. This ratio varies between 0.16 and 1.6, spanning a range of about a factor 10 (95% confidence interval). In this respect, the 2-D model predictions are slightly better than the 1-D predictions. The difference between predicted and mean acquisition range can be compensated for by introducing a single range correction factor, but this has no effect on the range of the confidence interval. It was shown that it is not possible to reduce this range by means of a simple modification of the existing model.
2-D ACQUIRE better predicts the effect of target orientation on acquisition performance, than the old version does. The analysis further showed, that the 2-D version of the model correctly predicts the slope of the probability *vs.* target range curve (averaged over a large number of conditions), and that the range correction factor is equal for identification and classification. This means that the relationship between the two levels in the present experiment with sea targets is correctly incorporated in the model. For 1-D ACQUIRE, there are minor differences between measured and predicted slope, and between the two range correction factors.
In conclusion, 2-D ACQUIRE may be used to predict *mean* acquisition performance for sea targets, if the predicted ranges are corrected by a factor 0.50. The 2-D version of ACQUIRE is preferable to the 1-D version. The model is not sophisticated enough to deal with individual cases.
These conclusions are based on a restricted image set, and only a number of aspects could be tested. With the observer performance data for simulated targets, TA models can be tested in more detail. For example, the effect of target orientation (both aspect angle and depression angle) is determined much more precisely, and also the effect of target contrast is measured. In Part 2 of this study (Bijl, 1996) it will be tested how adequate the 2-D model is in predicting these effects, and whether the factor 0.5 between predicted and measured acquisition ranges for sea targets is again found. Also the effect of interactions, e.g. the combined effect of target type and aspect angle or depression angle, can be determined. This effect is not modeled in ACQUIRE.

Acquisition of ground vs. sea targets

The results indicate that there are some important differences between acquisition of ground targets and sea targets. First, it seems to be relatively difficult to distinguish sea targets from one another. The ACQUIRE model, which is calibrated for ground targets, is far too

optimistic for sea targets. In Bijl and Valeton (1994), it was shown that for real FLIR imagery of ground targets (using the same sensor as in this study), ACQUIRE predictions are pessimistic rather than optimistic. Maybe, sea targets are more similar (contain fewer or less significant details) than ground targets relative to their dimensions. The ACQUIRE model only takes target dimensions, not the similarity of the targets in the set, as a measure.
A second finding is, that target acquisition performance decreases gradually with target range. Such a relationship is predicted by most TA models, but is not always found for approaching targets in a ground-to-ground target acquisition task. In the study of Bijl and Valeton (1992a), sometimes strong undulations in the relation between target distance and acquisition performance were found. These were ascribed to large local variations in background temperature which changed the thermal contrast and apparent shape of the target considerably. A sea background may be more uniform in temperature, which might facilitate modelling.
Finally, the uncertainty interval for the predictions for individual conditions is much larger than in the experiment with ground targets. In Bijl and Valeton (1994), the ratio between observed and predicted range varied between 0.9 and 3.6, i.e. a factor of 4. In the present study, the ratio spans a range of a factor 10. This difference may be due to differences in experimental conditions. In the ground-to-ground experiment, targets were only in front view. In the present air-to-sea experiment, aspect angle and depression angle were varied.

Differences between experiment and practical situation

In some aspects, the experimental conditions differ from the practical situation. For example, short sequences of the target approaches were presented in random order. In a practical situation, after a target is detected, the observer often sees a complete target approach in which he may accumulate information on the target. The reason for a random presentation order is, that it yields most information from the data. In Bijl and Valeton (1992a), image sequences were presented in both ways. It was found that the score at a certain distance for a complete target approach can be approximated quite well from the random order experiment by taking the maximum score of all image sequences between the largest and the actual distance, thus yielding a monotonic relationship between target range and observer performance. In the case that this relationship is already monotonic for a random presentation order, as in Fig. A2, there will be no difference in score for the two presentation orders. However, if there is a dip in acquisition performance, for example due to a rotation from side to front view during the approach as in Fig. A3, no dip is expected if the complete target approach is shown to the observer. Such dips are only found in a few of the 22 runs that were used in the experiment. The overall effect that is expected if complete target approaches are presented to the observer instead of random image sequences, is a slightly higher mean observer performance score, which comes a little closer to the prediction made by ACQUIRE.
A second difference with the practical situation in an ORION patrol aircraft is, that the observers had no distance information from the radar. This is especially important for the discrimination of the Tanker and the Coaster, which are very similar in shape, but different in size. Higher classification and recognition performance for these targets are expected if distance information is available. This might have consequences for the results of the validation of ACQUIRE. Therefore, the analysis was repeated using the data for the three

other targets only. The results for this set are quite similar to the results for the complete set.

ACKNOWLEDGMENTS

The author wishes to thank the observers from the Royal Dutch Navy ORION squadrons VSQ 320 and VSQ 321 (Marine Vliegkamp Valkenburg, Katwijk aan Zee) and LYNX helicopter squadron (Marine Vliegkamp de Kooy, Den Helder) for their helpful suggestions and for their contributions as observers in the experiment.

REFERENCES

Bijl, P. & Valeton, J.M. (1992a). *Observer experiments with BEST TWO thermal images. Part 2: Terrain interaction and Target motion* (Report IZF 1992 A-34). Soesterberg, The Netherlands: TNO Institute for Perception[2] (NATO Confidential).

Bijl, P. & Valeton, J.M. (1992b). *Observer experiments with BEST TWO thermal images. Part 3: Reliability of observer responses* (Report IZF 1992 A-35). Soesterberg, The Netherlands: TNO Institute for Perception (NATO Confidential).

Bijl, P. & Valeton, J.M. (1994). *Evaluation of target acquisition model "TARGAC" using "BEST TWO" observer performance data* (Report TNO-TM 1994 A-22). Soesterberg, The Netherlands: TNO Human Factors Research Institute.

Bijl, P. (1996). *Acquisition of sea targets. Part 2: Effect of target orientation and contrast on observer performance* (in preparation). Soesterberg, The Netherlands: TNO Human Factors Research Institute.

Johnson, J. (1958). Analysis of image forming systems. In *Image Intensifier Symposium,* Fort Belvoir, VA: Warfare Vision Branch, Electrical Engineering Dept., US Army Engineer Research and Development Laboratories.

De Jong, A.N. (1994). *NL Ship-IR imagery at MAPTIP* (Rep. FEL-94-A143). The Hague, The Netherlands: TNO Physics and Electronics Laboratory.

De Jong, A.N., Janssen, Y.H.L., Roos, M.J.J. & Kemp, R.A.W. (1991). *Infrared and Electro-Optical experiments during BEST TWO by Research Group Infrared* (Rep. FEL-91-A252). The Hague, The Netherlands: TNO Physics and Electronics Laboratory (NATO Restricted).

Luria, S.M., Kinney, J.A.S., Schlichting, C.L. & Ryan, A.P. (1979). *The limiting effects of astigmatism on visual performance through periscopes* (Report 905). Groton, CT: Naval Submarine Medical Research Laboratory, Naval Submarine Base.

Obert, L.P. *et al.* (1990). An experimental Study of the effect of vertical resolution on FLIR performance. *IRIS Passive Sensors Imaging Symposium.*

Ratches, J.A. (1976). Static Performance Model for Thermal Imaging Systems. *Optical Engineering 15, 6,* 525-530.

Ratches, J.A., Lawson, W.R., Shields, F.J., Hoover, C.W., Obert, L.P., Rodak. S.P. & Sola, M.C. (1981). *Status of Sensor Performance Modelling at NV&EOL.* Fort Belvoir, VA: Night Vision & Electro-Optics Laboratory.

Scott, L. (1990). *C2NVEO Thermal Imaging Systems Performance Model - FLIR90 version 0* (Report R5008986). Fort Belvoir, VA: Center for Night Vision & Electro-Optics.

Scott, L. (1990). *C2NVEO Model - ACQUIRE version 0* (Report R5008988). Fort Belvoir, VA: Center for Night Vision & Electro-Optics.

Valeton, J.M. & Bijl, P. (1992). *Observer experiments with BEST TWO thermal images. Part 1: Design, training and Observer selection* (Report IZF 1992 A-33). Soesterberg, The Netherlands: TNO Institute for Perception (NATO Confidential).

Valeton, J.M. & Bijl, P. (1994). *Target Acquisition: Human Observer Performance Studies and TARGAC model validation* (Report TNO-TM 1994 B-15). Soesterberg, The Netherlands: TNO Human Factors Research Institute.

---

[2] On January 1, 1994 the name "TNO Institute for Perception" has been changed to "TNO Human Factors Research Institute".

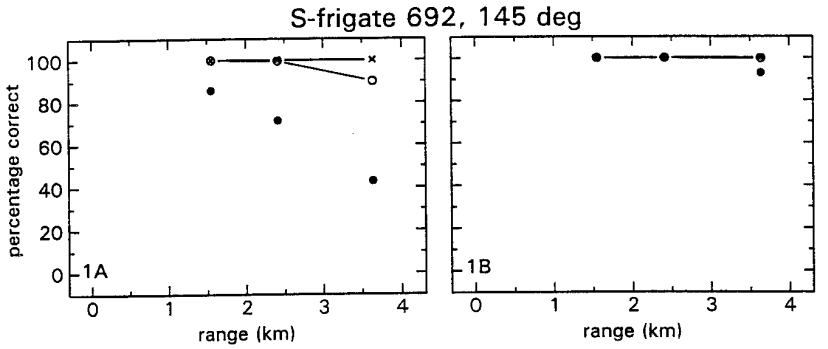
Walraven, J., Boogaard, J. & Alferdinck, J.W.A.M. (1995). *Een visuskaart voor de bepaling van de gezichtsscherpte op de kijkafstand van beeldschermen* [A test for measuring visual acuity at the viewing distance of visual display units] (Report TNO-TM 1995 C-10). Soesterberg, The Netherlands: TNO Human Factors Research Institute.

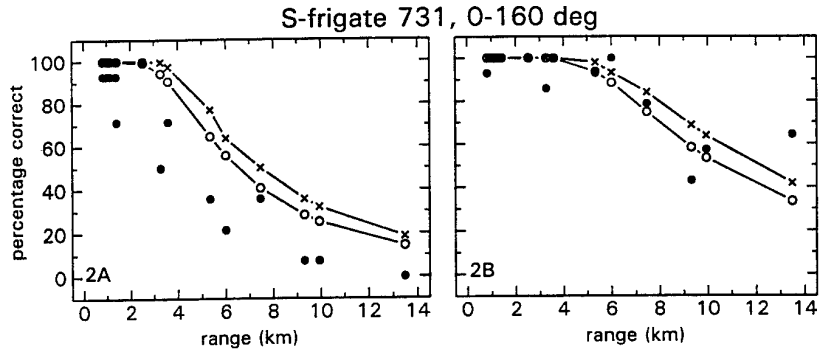
Soesterberg, 17 September 1996

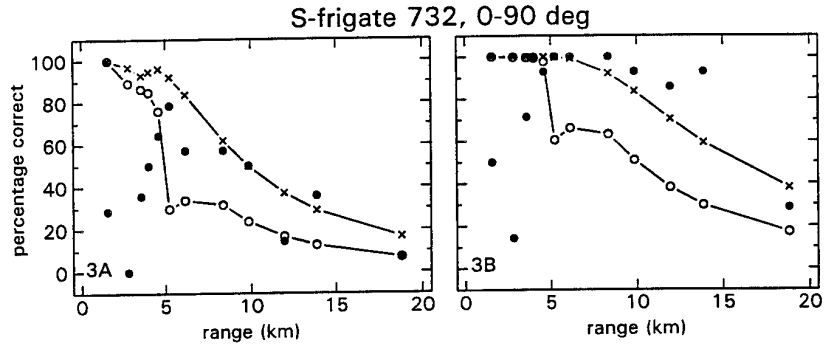
Dr. P. Bijl
(author, project manager)

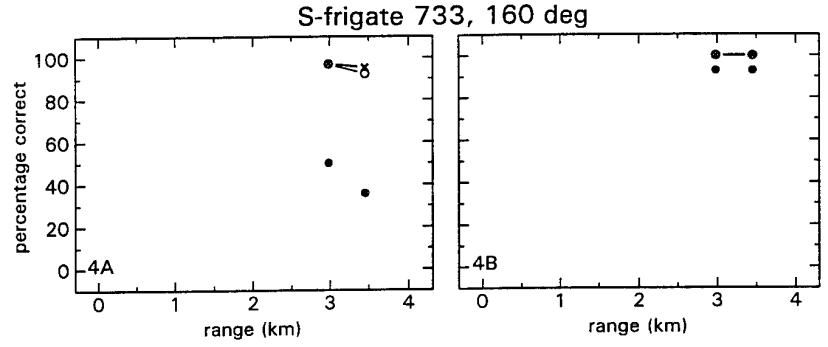
APPENDIX I Complete set of observer performance data and ACQUIRE predictions

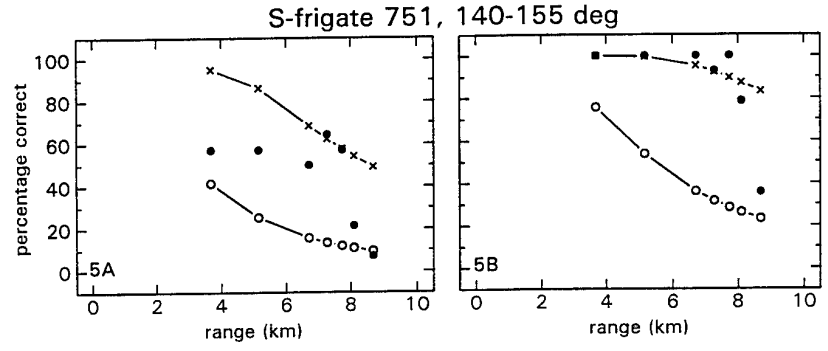
In Fig. A1-A22, the complete set of observer performance data and ACQUIRE predictions are presented. The identification predictions are presented in Figs A, and the classification results in Figs B. In each plot, the target type is given, followed by a number (which is a composition of the original FEL tape-number and a sequence number). The observer data are indicated by filled circles. Open circles indicate the 1-D ACQUIRE calculations, crosses the 2-D predictions. In most runs, target range was the main variable, although the aspect angle usually varied as well. In that case, acquisition probability (in %) is plotted as a function of target range (in km), and the aspect angle of the ship, or a range of aspect angles, is indicated. In the case of a circle run (e.g. Figs A19 and A20) probability is plotted as a function of aspect angle (in degrees), and target range is indicated.

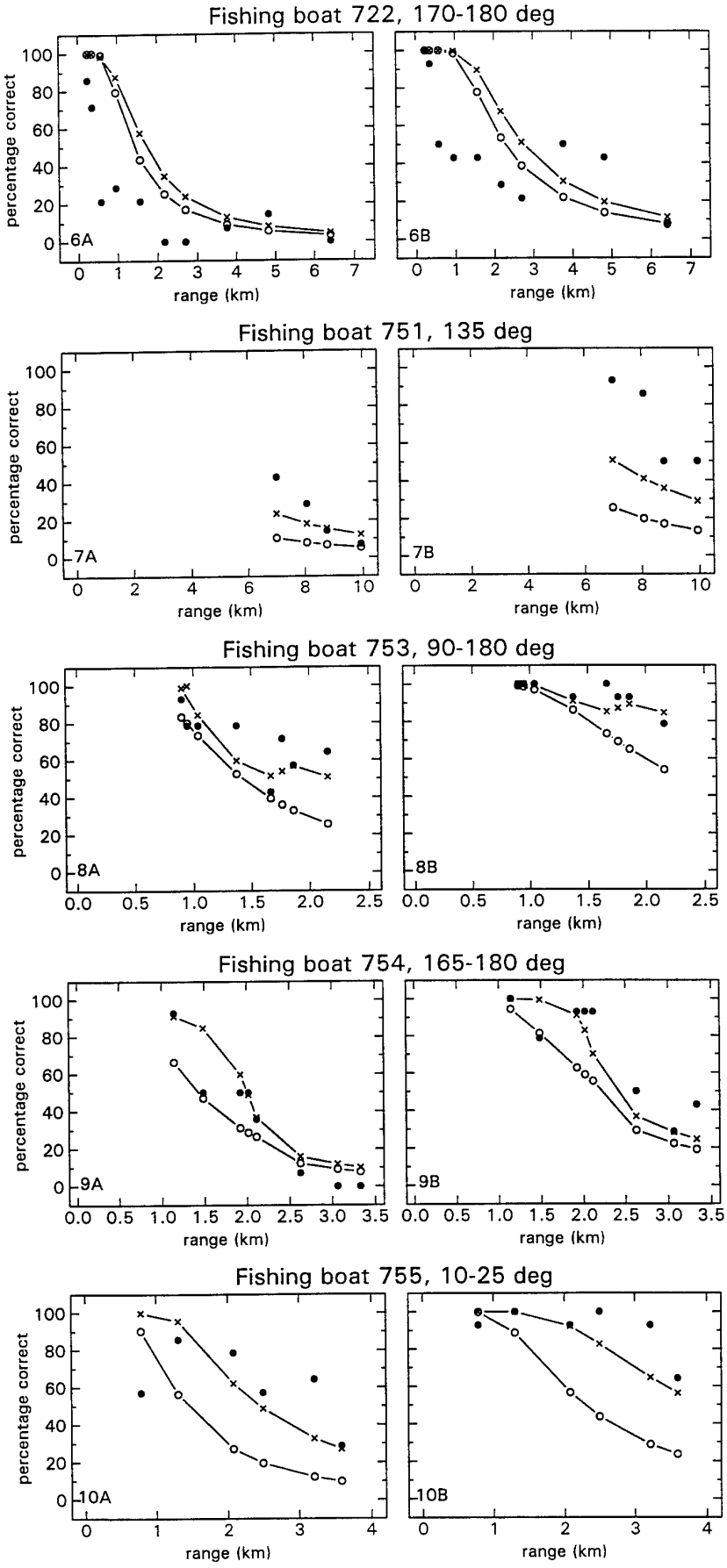

S-frigate 692, 145 deg
percentage correct
range (km)
1A
1B

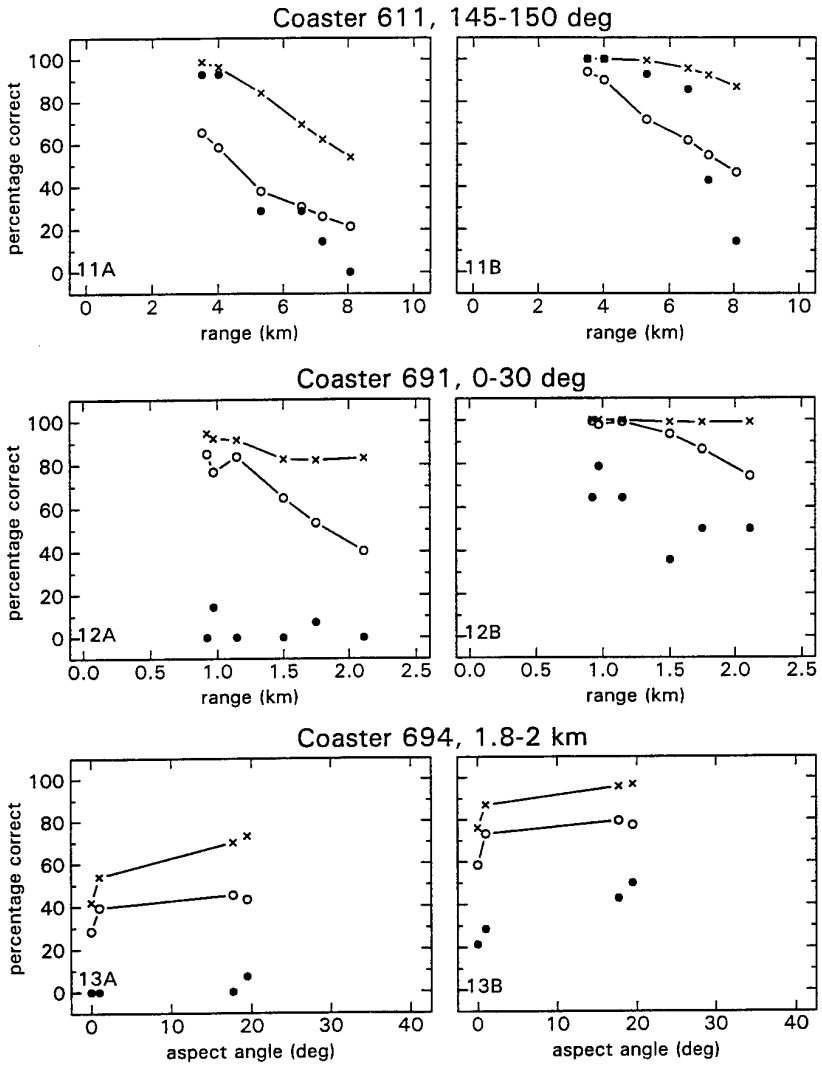

S-frigate 731, 0-160 deg
percentage correct
range (km)
2A
2B

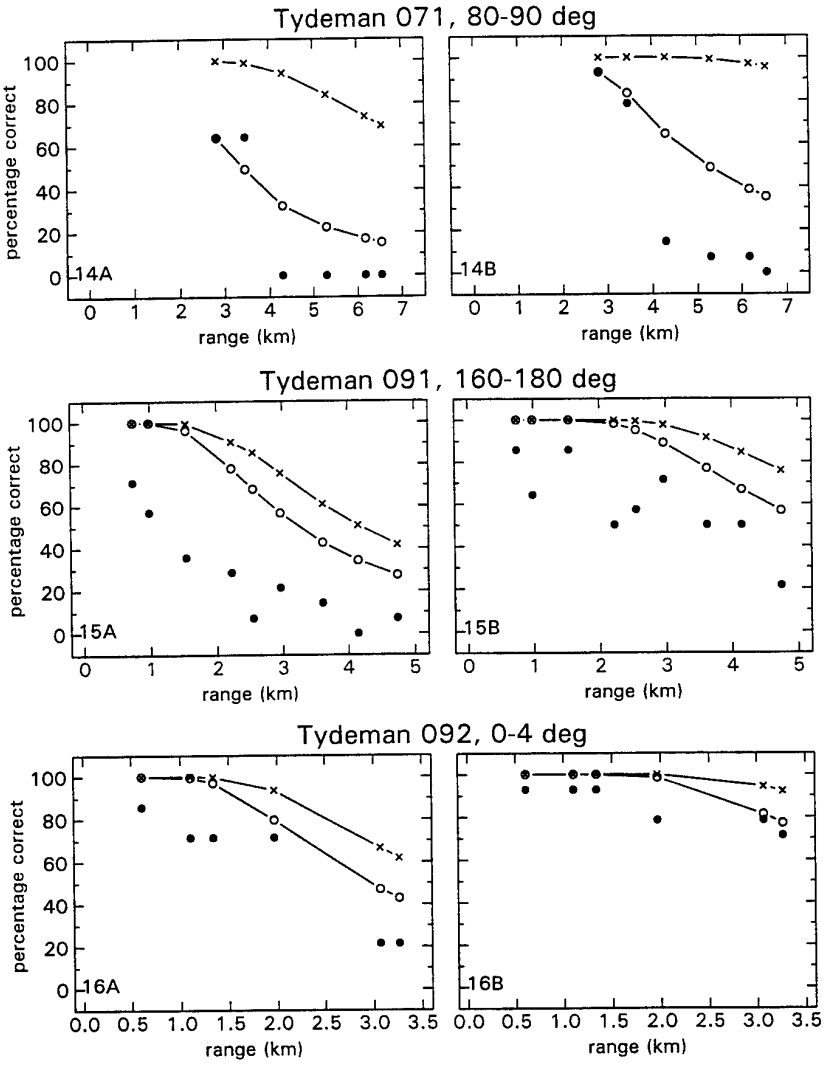

S-frigate 732, 0-90 deg
percentage correct
range (km)
3A
3B

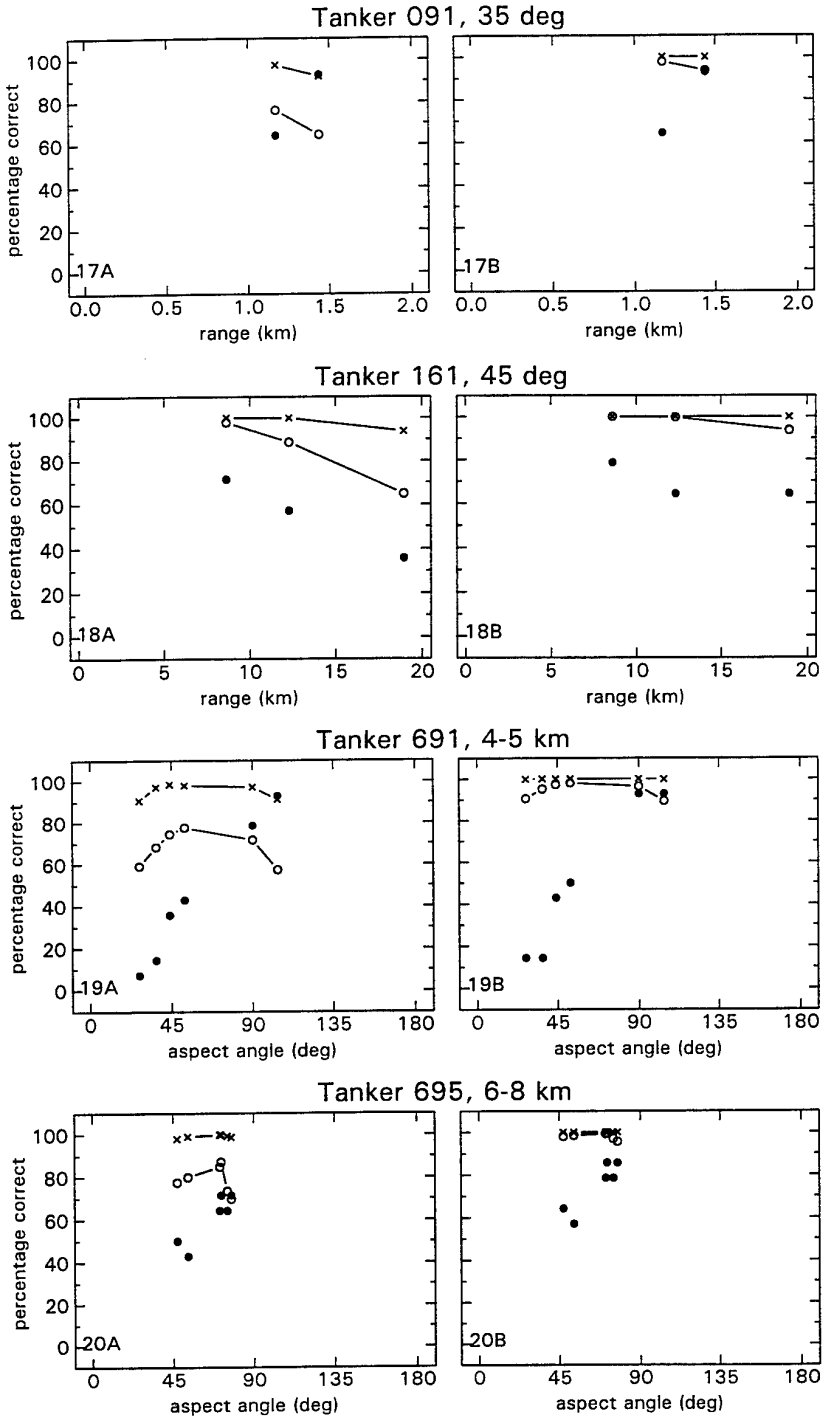

S-frigate 733, 160 deg
percentage correct
range (km)
4A
4B

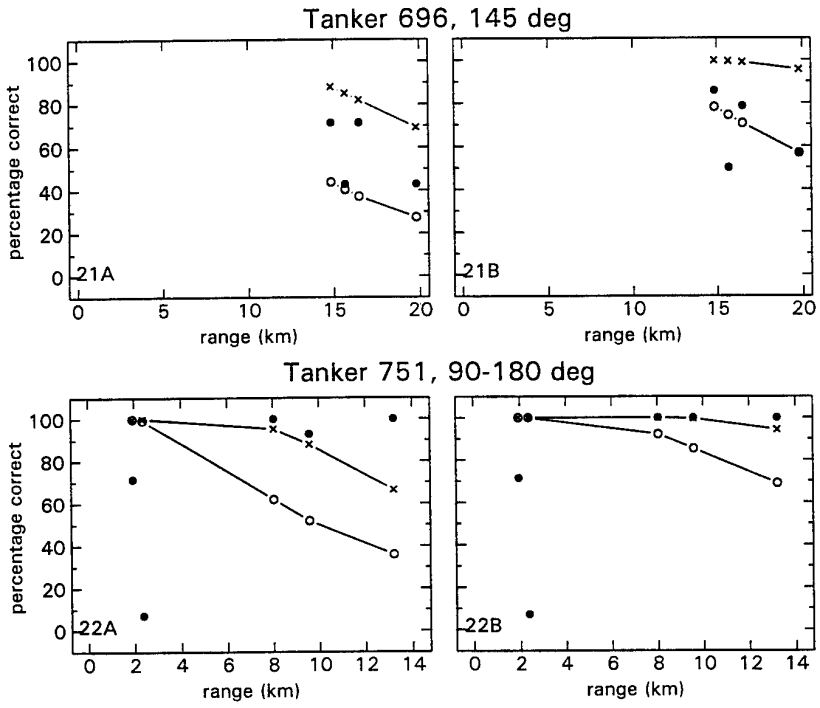

S-frigate 751, 140-155 deg
percentage correct
range (km)
5A
5B


Fishing boat 722, 170-180 deg
percentage correct
range (km)
6A
6B
Fishing boat 751, 135 deg
7A
7B
Fishing boat 753, 90-180 deg
8A
8B
Fishing boat 754, 165-180 deg
9A
9B
Fishing boat 755, 10-25 deg
10A
10B


Coaster 611, 145-150 deg
percentage correct
11A
11B
range (km)
Coaster 691, 0-30 deg
12A
12B
range (km)
Coaster 694, 1.8-2 km
13A
13B
aspect angle (deg)


Tydeman 071, 80-90 deg
percentage correct
14A
14B
range (km)
Tydeman 091, 160-180 deg
15A
15B
range (km)
Tydeman 092, 0-4 deg
16A
16B
range (km)


Tanker 091, 35 deg
percentage correct
17A
17B
range (km)
Tanker 161, 45 deg
18A
18B
range (km)
Tanker 691, 4-5 km
19A
19B
aspect angle (deg)
Tanker 695, 6-8 km
20A
20B
aspect angle (deg)


Tanker 696, 145 deg
percentage correct
100
80
60
40
20
0
21A
0
5
10
15
20
range (km)
21B
0
5
10
15
20
range (km)
Tanker 751, 90-180 deg
percentage correct
100
80
60
40
20
0
22A
0
2
4
6
8
10
12
14
range (km)
22B
0
2
4
6
8
10
12
14
range (km)

APPENDIX II Derivation of some rules of thumb

In Figs A23–A27, the observer data for each target type are plotted as a function of (target area)$^{-1/2}$ and a rough overall curve is drawn through the data to estimate the 50% correct (target area)$^{1/2}$ (in mrad). The results are used to derive some estimates of target acquisition ranges from the data (see § 5.4). Estimates of the 50% correct target square-root areas are given in Table A1.

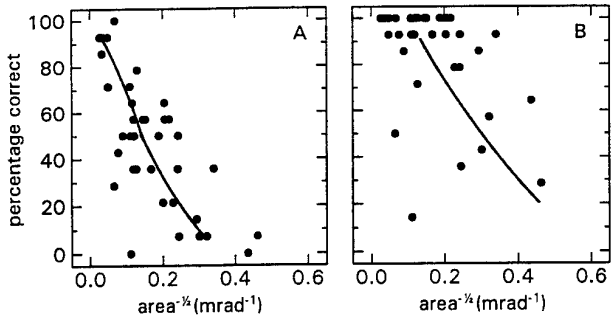


Fig. A23 Entire set of observer scores for the S-frigate as a function of (target area)$^{-1/2}$. A: Identification, B: Classification. A rough overall curve was drawn through the data to estimate the 50% correct target square-root area.

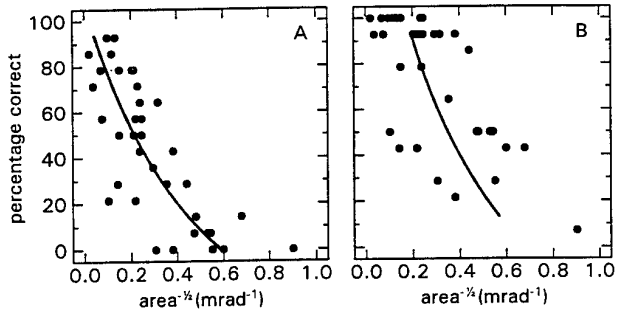


Fig. A24 Entire set of observer scores for Fishing Boats as a function of (target area)$^{-1/2}$. See also Fig. A23.

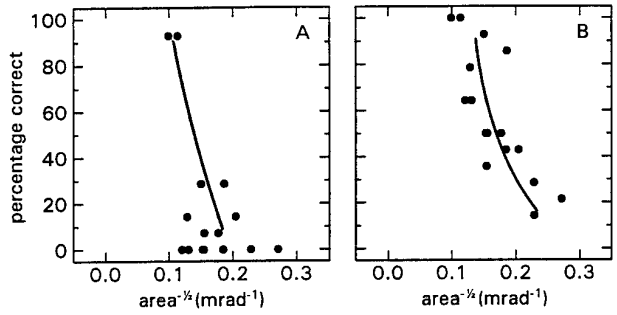


Fig. A25 Entire set of observer scores for Coasters as a function of (target area)$^{-1/2}$. See also Fig. A23.

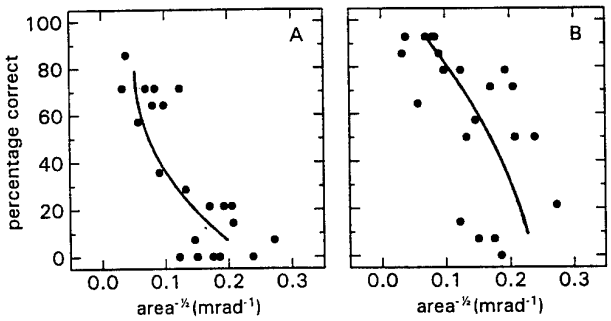


Fig. A26 Entire set of observer scores for the Tydeman as a function of (target area)$^{-1/2}$. See also Fig. A23.

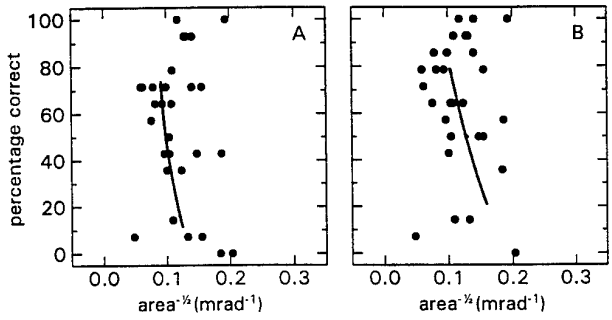


Fig. A27 Entire set of observer scores for Tankers as a function of (target area)$^{-1/2}$. See also Fig. A23.

Table A1 Estimates of the 50% correct target square-root areas for each target separately. Column 2: identification, column 3: classification.

target type	50% identification area$^{1/2}$ (mrad)	50% classification area$^{1/2}$ (mrad)
S-frigate	7	3
Fishing boat	4	3
Coaster	7	6
Tydeman	11	6
Tanker	10	8

REPORT DOCUMENTATION PAGE

1. DEFENCE REPORT NUMBER (MOD-NL)	2. RECIPIENT'S ACCESSION NUMBER	3. PERFORMING ORGANIZATION REPORT NUMBER
RP 96-0177		TM-96-A037
4. PROJECT/TASK/WORK UNIT NO.	**5. CONTRACT NUMBER**	**6. REPORT DATE**
786.1	A94/KM/315	17 September 1996
7. NUMBER OF PAGES	**8. NUMBER OF REFERENCES**	**9. TYPE OF REPORT AND DATES COVERED**
47	16	Final

10. TITLE AND SUBTITLE

Acquisition of sea targets. Part 1: Observer performance and "ACQUIRE" model predictions for air-to-surface FLIR imagery

11. AUTHOR(S)

P. Bijl

12. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

TNO Human Factors Research Institute
Kampweg 5
3769 DE SOESTERBERG

13. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Director of Navy Research and Development
P.O. Box 20702
2597 PC DEN HAAG

14. SUPPLEMENTARY NOTES

15. ABSTRACT (MAXIMUM 200 WORDS, 1044 BYTE)

In this study, two laboratory experiments were carried out to test how well human observers using an electro-optical (E/O) viewing device, are able to identify or classify sea targets. Such knowledge is of interest to evaluate the applicability of target acquisition (TA) models for a sea environment. Most models are designed and tested for ground targets and backgrounds, and their reliability for acquisition of sea targets is unknown. In the first experiment, which is reported here, observer performance was measured on real thermal infrared (FLIR) imagery that was collected on ORION flights. Experienced observers from several Royal Dutch Navy ORION and LYNX helicopter squadrons participated in this experiment. First, some rules-of-thumb were deduced from the data. For example, with the sensor that was used in the experiment and good atmospheric conditions, an S-frigate in side view may be classified (50% correct) at 14 km and identified at 7 km. For the same target in front view these ranges are about 5 km and 2 km, respectively. For the Tydeman and a Fishing Boat in side view, the 50% correct classification and identification ranges are about 7 and 4 km. Second, the results of the experiment were used to test a widely used TA model, ACQUIRE. Important differences were found between measured and predicted performance. On average, the most recent version of ACQUIRE overestimates acquisition ranges by a factor of 2. Further, the model does not give accurate predictions for individual situations: the ratio between measured and predicted acquisition range depends largely on the circumstances and varies between 0.16 and 1.6 (95%-criterion). The relation between acquisition probability and target range, and the ratio between identification and classification range, are correctly predicted by the model. Apparently, these relations are the same for sea and ground targets. The most recent version of ACQUIRE is preferable to the old version, which is still widely used. The model may be used to predict mean acquisition performance for sea targets, if the predicted ranges are corrected by a factor 0.50. In the second, more extensive experiment, the applicability of models for acquisition of sea targets are tested in more detail. The results will be reported in Part 2 of this study.

16. DESCRIPTORS	IDENTIFIERS
Infrared Model Evaluation Navy Observer Performance Target Acquisition	Air-to-Surface

17a. SECURITY CLASSIFICATION (OF REPORT)	17b. SECURITY CLASSIFICATION (OF PAGE)	17c. SECURITY CLASSIFICATION (OF ABSTRACT)
18. DISTRIBUTION/AVAILABILITY STATEMENT		17d. SECURITY CLASSIFICATION (OF TITLES)
Unlimited availability		

VERZENDLIJST

1.	Directeur M&P DO
2.	Directie Wetenschappelijk Onderzoek en Ontwikkeling Defensie
3. {	Hoofd Wetenschappelijk Onderzoek KL
	Plv. Hoofd Wetenschappelijk Onderzoek KL
4.	Hoofd Wetenschappelijk Onderzoek KLu
5. {	Hoofd Wetenschappelijk Onderzoek KM
	Plv. Hoofd Wetenschappelijk Onderzoek KM
6, 7 en 8.	Bibliotheek KMA, Breda
9 tm 13.	KLTZ F.S. Vleer, MARSTAF/LUVRT, Den Haag
14 en 15.	Dr. J.S. de Vries, TNO-FEL, Afdeling Elektro-Optiek, Den Haag

Extra exemplaren van dit rapport kunnen worden aangevraagd door tussenkomst van de HWOs of de DWOO.